METHOD AND SYSTEM FOR PREDICTING NUCLEIC ACID HYBRIDIZATION THERMODYNAMICS AND COMPUTER-READABLE STORAGE MEDIUM FOR USE THEREIN

CROSS-REFERENCE TO RELATED APPLICATION

This application claims the benefit of U.S. provisional application Serial No. 60/209,778, filed June 7, 2000, entitled "Nucleic Acid Hybridization Prediction and Biochemical Techniques Utilizing Same."

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

This invention was made with government support under Grant No. HG02020 provided by NIH. The United States Government has certain rights in the invention.

BACKGROUND OF THE INVENTION

1. Field of the Invention

This invention relates to methods and systems for predicting nucleic acid hybridization thermodynamics and computer-readable storage medium for use therein.

2. Background Art

Improvement of the efficiency of hybridization-based techniques requires the optimization of the binding between two sequences. Accurate prediction of the thermodynamics allows optimal choice of the sequences, temperature, and salt conditions. Hence, the prediction of nucleic acid thermodynamics is important to optimize techniques like PCR (Saiki et al., 1988), Southern and Northern blotting (Southern, 1975), antigene targeting (Freier, 1993), and Kunkel site-directed mutagenesis (Kunkel et al., 1987).

Hybridization prediction is also important for designing DNA microchips that have a wide field of application ranging from diagnostics (Hacia, 1999; Yershov et al., 1996) to gene expression analysis (Ferea et al., 1999) and drug discovery (Debouk and Goodfellow, 1999). Microchips contain a large number of DNA probe sequences that have to be designed to specifically hybridize target sequences in a pool of DNA fragments. First, a DNA probe should be designed to bind to only one site of only one DNA target. Second, the different DNA probe sequences need to hybridize to their targets under the same temperature and solution conditions. Moreover, in sequencing by hybridization (Fodor et al., 1993; Mirzabekov, 1994) where microchips are used to determine the sequence of given DNA, one has to be able to know hybridization thermodynamics to discriminate signals resulting from perfectly matched and mismatched probe/target hybridizations.

Another widely used technique that requires hybridization prediction is the fluorescence *in situ* hybridization (FISH) technique (Gall and Pardue, 1969). In this technique, a fluorescent tagged nucleic acid probe is designed to specifically hybridize cellular or tissue section nucleic acids. The target of these probes can either be endogenous DNA, messenger RNA or viral and bacterial sequences.

Therefore, FISH is used to monitor gene expression (McNicol and Farquharson, 1997), detect infectious agents (Bashir et al., 1994; McNicol and Farquharson, 1997; Pollanen et al., 1993), study cell cycle (McNicol and Farquharson, 1997), map chromosomes and study nuclear architecture (Heng et al., 1997). It was also determined that a set of probes can be used simultaneously (multiFISH) to detect different loci (Pagon, 1997). Once again, prediction of hybridization is essential to insure specificity. Nucleic-acid hybridization prediction is also important for the design of oligonucleotide aptamers or antisense oligonucleotides (Cohen, 1992) that can be used for various therapeutic applications. A new type of probes known as molecular beacons (Bonnet et al., 1999; Tyagi et al. 1998) that are very specific has been developed and shown to be efficient for mutation analysis (Giensendorf et al., 1998) and multiplex detection of single nucleotide variations (Marras et al., 1999). The design and prediction of the

thermodynamics of these beacons is helped by hybridization thermodynamics prediction (Bonnet et al., 1999). Accurate prediction of hybridization is also important for the practical realization of DNA-based or more generally nucleic acid-based computers. (Adleman, L.M., 1994).

The development of molecular biology techniques based on hybridization (PCR, FISH, DNA microchips, etc.) has resulted in a need for efficient automated ways to design probes and primers. In the last decade, numerous algorithms have been developed to optimize the design of primers and probes for various applications (Rychlik and Rhoads, 1989)(Breslauer et al., 1986; Chen and Zhu, 1997; Dopazo et al., 1993; Haas et al., 1998; Hillier and Green, 1991; Hyndman et al., 1996; Li et al., 1997; Link et al., 1997; Pesole et al., 1998; Proutski and Holmes, 1996). Numerous unpublished software to predict primers are also made available by research groups and biotech companies on the World Wide Web (Primer3 from the Whitehead Institute for Biomedical Research. Primer Express™ from PE Biosystems, DNAstar from IDT, etc.)

There are currently many software packages on the market for DNA primer design including: OLIGO, PRIMER PREMIER, OSP, GCG, PrimerMaster, and Primo. None of the current programs, however, were written by experts in DNA thermodynamics; thus, there are many improvements that can be made. Nearly all of the current software packages contain mistakes that result from a lack of understanding of the underlying theory of DNA hybridization. PCR is a fairly robust process and thus even crude programs make predictions that work 90-95% of the time. Multiplex PCR primer design, however, is not at all trivial and detailed knowledge of the physical chemistry of DNA hybridization, and the availability of an accurate thermodynamic database are essential to reliable design of multiplex PCR primers. In multiplex PCR, several primers must be designed to specifically bind to different sites on target DNA at a given temperature with minimal background binding to mismatch sites and with minimal cross-hybridizations between pairs of primers. The design of molecular beacons for DNA oligonucleotide arrays is also very challenging because of the complex competing equilibria.

Most of the existing programs aiming at finding an optimum probe that binds a specific location on a target, however, do not include accurate stability rules for hybridization and neglect or poorly approximate competitive binding sites, strand folding and strand dimerization.

U.S. Patent Nos. 5,593,834 and 6,027,884 to Lane et al. disclose methods to design and construct DNA sequences with selected reaction attributes.

In summary, prediction of nucleic acids thermodynamics is important to optimize various molecular biology techniques including multiplex PCR, DNA microchips, molecular beacons, and fluorescence *in situ* hybridization. Most of the available programs for probe design do not include a complete parameterization and often do not account for mismatches. Moreover, single strand folding is not taken into account, which often leads to inaccurate predictions.

SUMMARY OF THE INVENTION

An object of the invention is to provide a method and system for predicting nucleic acid hybridization thermodynamics and computer-readable storage medium for use therein wherein the invention utilizes a thermodynamically rigorous approach to evaluate the quality of probes and simulate probe/target hybridization.

Another object of the invention is to provide a method and system for predicting nucleic acid hybridization thermodynamics and computer-readable storage medium for use therein wherein the invention also takes into account single strand folding thermodynamics to calculate effective hybridization thermodynamics.

In carrying out the above objects and other objects of the present invention, a method for predicting nucleic acid hybridization thermodynamics is provided. The method includes providing a database of thermodynamic parameters, receiving hybridization information which represents at least one sequence, receiving correction data, receiving a first set of data which represents hybridization conditions, and calculating hybridization thermodynamics including net

hybridization thermodynamics based on the hybridization information, the thermodynamic parameters, the correction data and the first set of data.

The hybridization thermodynamics of individual single stranded, bimolecular and higher order complexes may be statistically weighted in a numerical process and the equilibrium concentration of each species is output.

The correction data may include folding correction data and/or linear correction data.

The thermodynamic parameters may include DNA thermodynamic parameters.

The DNA thermodynamic parameters may include dangling end parameters and/or coaxial stacking parameters.

The DNA thermodynamic parameters may further include terminal mismatch parameters.

The thermodynamic parameters may include RNA thermodynamic parameters and/or hybrid DNA/RNA thermodynamic parameters.

The thermodynamic parameters may further include DNA loop thermodynamic parameters.

The hybridization information may represent top and bottom strand sequences which form a duplex and wherein the hybridization thermodynamics are calculated for the duplex.

The hybridization information may further represent at least a section of a target and a length of at least one primer or probe complimentary to the target.

The hybridization thermodynamics may be calculated for a plurality of primers or probes complimentary to the target.

The hybridization information may represents at least a section of a target and a primer or probe.

A length of the target may be longer than a length of the primer or probe and wherein the hybridization thermodynamics are calculated for a best target/primer or target/probe complex and for competitive mismatch complexes.

Hybridization information may represent at least a section of a target and a primer or probe and wherein a length of a target is longer than the length of the primer or probe and wherein the hybridization thermodynamics are calculated for a best target/primer or target/probe complex and for competitive target/primer or target/probe complexes.

The method may further include calculating concentration of each species in a solution at a plurality of temperatures.

Hybridization information may also represent a primer or probe and wherein the length of the target is longer than a length of the primer or probe and wherein the hybridization thermodynamics are calculated for a best target/primer or target/probe complex and for competitive mismatch complexes and wherein the method may further comprise calculating concentration of every species in a solution at a plurality of temperatures.

The hybridization thermodynamics may be calculated for at least two best target/primer or target/probe complexes and for their corresponding competitive mismatch complexes and wherein the method may further comprise correcting for any interactions between the at least two best target/primer or target/probe complexes and their components.

Further in carrying out the above objects and other objects of the present invention, a system for predicting nucleic acid hybridization thermodynamics is provided. The system includes a database of thermodynamics parameters, means for receiving hybridization information which represents at least one sequence, and means for receiving correction data. The system further includes receiving a first set of data which represents hybridization conditions, and means for calculating hybridization thermodynamics including net hybridization thermodynamics based on the hybridization information, the thermodynamic parameters, the correction data and the first set of data.

The hybridization thermodynamics of individual single stranded, bimolecular and higher order complexes may be statistically weighted in a numerical process and the equilibrium concentration of each species is output.

The correction data may include folding correction data and/or linear correction data.

The thermodynamic parameters may include DNA thermodynamic parameters such as dangling end parameters.

The DNA thermodynamic parameters may include coaxial stacking parameters and/or terminal mismatch parameters.

The thermodynamic parameters may include RNA thermodynamic parameters and/or hybrid DNA/RNA thermodynamic parameters.

The thermodynamic parameters may further include DNA loop thermodynamic parameters.

The hybridization information may represent top and bottom strand sequences which form a duplex and wherein the hybridization thermodynamics are calculated for the duplex.

The hybridization information may also represent at least a section of a target and a length of at least one primer or probe complimentary to the target.

The hybridization thermodynamics may be calculated for a plurality of primers or probes complimentary to the target.

The hybridization information may represent at least a section of a target and a primer or probe.

A length of the target may be longer than a length of the primer or probe and wherein the hybridization thermodynamics are calculated for a best target/primer or target/probe complex and for competitive mismatch complexes.

Hybridization information may represent at least a section of a target and a primer or probe and wherein a length of a target is longer than the length of the primer or probe and wherein the hybridization thermodynamics are calculated for a best target/primer or target/probe complex and for competitive target/primer or target/probe complexes.

The system may further include means for calculating concentration of each species in a solution at a plurality of temperatures.

Hybridization information may also represent a primer or probe and wherein the length of the target is longer than a length of the primer or probe and wherein the hybridization thermodynamics are calculated for a best target/primer or target/probe complex and for competitive mismatch complexes and wherein the system may further comprise means for calculating concentration of every species in a solution at a plurality of temperatures.

The hybridization thermodynamics may be calculated for at least two best target/primer or target/probe complexes and for their corresponding competitive mismatch complexes and wherein the system may further comprise

means for correcting for any interactions between the at least two best target/primer or target/probe complexes and their components.

Still further in carrying out the above objects and other objects of the present invention, a computer-readable storage medium having stored therein a database of thermodynamics parameters and a computer program are provided. The computer program executes the steps of: a) receiving hybridization information which represents at least one sequence; b) receiving correction data; c) receiving a first set of data which represents hybridization conditions; and d) calculating hybridization thermodynamics based including net hybridization thermodynamics based on the hybridization information, the thermodynamic parameters, the correction data and the first set of data.

The hybridization thermodynamics of individual single stranded, bimolecular and higher order complexes may be statistically weighted in a numerical process and the equilibrium concentration of each species is output.

The correction data may include folding correction data and/or linear correction data.

The thermodynamic parameters may include DNA thermodynamic parameters.

The DNA thermodynamic parameters may include dangling end parameters and/or coaxial stacking parameters.

The DNA thermodynamic parameters may further include terminal mismatch parameters.

The thermodynamic parameters may include RNA thermodynamic parameters and/or hybrid DNA/RNA thermodynamic parameters.

The thermodynamic parameters may further include DNA loop thermodynamic parameters.

The hybridization information may represent top and bottom strand sequences which form a duplex and wherein the hybridization thermodynamics are calculated for the duplex.

The hybridization information may represent at least a section of a target and a length of at least one primer or probe complimentary to the target.

The hybridization thermodynamics may be calculated for a plurality of primers or probes complimentary to the target.

The hybridization information may represent at least a section of a target and a primer or probe.

A length of the target may be longer than a length of the primer or probe and wherein the hybridization thermodynamics are calculated for a best target/primer or target/probe complex and for competitive mismatch complexes.

Hybridization information may represent at least a section of a target and a primer or probe and wherein a length of a target is longer than the length of the primer or probe and wherein the hybridization thermodynamics are calculated for a best target/primer or target/probe complex and for competitive target/primer or target/probe complexes.

The program may further execute the step of calculating concentration of each species in a solution at a plurality of temperatures.

Hybridization information may also represent a primer or probe and wherein the length of the target is longer than a length of the primer or probe and wherein the hybridization thermodynamics are calculated for a best target/primer or target/probe complex and for competitive mismatch complexes and wherein the

program may execute the step of calculating concentration of every species in a solution at a plurality of temperatures.

The hybridization thermodynamics may be calculated for at least two best target/primer or target/probe complexes and for their corresponding competitive mismatch complexes and wherein the program may execute the step of correcting for any interactions between the at least two best target/primer or target/probe complexes and their components.

The above objects and other objects, features, and advantages of the present invention are readily apparent from the following detailed description of the best mode for carrying out the invention when taken in connection with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGURE 1 is a schematic drawing wherein multiple equilibria are considered for concentration calculations;

FIGURE 2a is a schematic drawing of a user input interface wherein the user provides various input information for a first module of the invention;

FIGURE 2b is a schematic drawing of a user output interface wherein a computer provides output information corresponding to the input information of Figure 2a;

FIGURE 3a is a schematic drawing of a user input interface wherein the user provides various input information for a second module of the invention;

FIGURE 3b is a schematic drawing of a user output interface wherein a computer provides output information corresponding to the input information of Figure 3a;

FIGURE 4a is a schematic drawing of a user input interface wherein the user provides various input information for a third module of the invention;

FIGURE 4b is a schematic drawing of a user output interface wherein a computer provides output information corresponding to the input information of Figure 4a;

FIGURE 5a is a schematic drawing of a user input interface wherein the user provides various input information for a fifth module of the invention;

FIGURE 5b is a schematic drawing of a user output interface wherein a computer provides output information corresponding to the input information of Figure 5a;

FIGURE 6 is a block diagram flow chart illustrating the solution of conservation equations of the present invention;

FIGURE 7 is a schematic diagram illustrating multiplex PCR design;

FIGURE 8 shows prediction of molecular beacon net hybridization thermodynamics;

FIGURE 9 shows simulation of molecular beacon hybridization concentrations at temperatures from 0 to 100°C;

FIGURE 10 is a diagram of match vs. mismatch hybridization;

FIGURE 11 shows match vs. mismatch hybridization simulation at different temperatures;

FIGURE 12 shows a general case of competitive hybridization equilibria that can be solved using the described numerical methods; and

FIGURE 13 is an example of simultaneous equations for the general five molecule case.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

In general, the method and system of the present invention include rigorous thermodynamic parameterization for Watson-Crick base pairs, internal mismatches, terminal mismatches, terminal dangling ends, co-axial stacking interactions, sodium and magnesium salt dependence, denaturants (urea, formamide, DMSO). In addition, loop parameters for hairpins, internal loops, bulges, and multibranched loops are included. For DNA essentially all the parameters have been previously published or all included in the Appendix hereto. Specifically, the parameters which have been published include Watson-Crick parameters, sodium dependence, GT, GA, CT, AC, AA, CC, GG, and TT mismatches. The parameters included herein include dangling ends, terminal mismatches, DNA loop parameters, and co-axial stacking parameters. For RNA, the parameters have been published by Douglas H. Turner et al. For DNA/RNA hybrid duplexes, the parameters have been published by Naoki Sugimoto.

The method and system are adapted for future implementation of parameters for modified nucleosides (including but not limited to inosine, 5-nitroindole, PNA, MOE-modified RNA, and iso-bases). With these parameters, it is possible to predict the melting temperature, Tm, of a duplex within 2°C on average. Correction for surface effects for DNA chip arrays is also implemented. In addition to predicting duplex hybridization, the software accounts for single-strand secondary structure. This is accomplished by a new numerical procedure for solving complex coupled equilibria (multi-state model). With this approach, it is possible to accurately predict not only the Tm for hybridization but also the concentration of every species in the solution (*e.g.* match duplex, mismatch, duplex, folded target, folded primer, primer dimer, etc.) at every temperature from 0 to 100°C. Thus, it is possible to use this software to design oligonucleotide hybridization with optimized temperature, salt, and strand concentrations.

Predicting Accurately Primer Target Interaction Stability

The stability of a primer/target or probe/target complex can be described by the free energy of association of the probe and the target. The most accurate way to calculate free energy of association is to use the nearest-neighbor model with accurate thermodynamic parameters. Thermodynamic parameters should account for Watson-Crick base pairs (SantaLucia et al., 1996; Allawi & SantaLucia, 1997; SantaLucia, 1998), single mismatches (Allawi & SantaLucia, 1997), terminal mismatches (disclosed herein), dangling ends (Bommarito, Pugret & SantaLucia, 2000) and possibly double mismatches. Proper calculation of the monovalent and divalent salt dependence is also important (SantaLucia, 1998). Other loop motifs for hairpins, bulge, internal loops and multi-branched loops are important for single strand secondary structure prediction, but are often very crudely approximated. Moreover, when primer and target folding can occur, a set of coupled equilibria should be used to model the system. The nearest-neighbor model needs to be used to determine the equilibrium constant of each equilibrium. The determination of possible primer or target folding can be addressed by using secondary-structure prediction algorithms like M. Zuker's MFOLD (Zuker, 1989).

Secondary Structure and Net Hybridization Thermodynamics

Species Concentration Calculations

Consider a system of strands S1 and S2 with four states: folded target, folded probe, probe bound to target, and random coil target and probe. The model can be described by three equilibria as shown in Figure 1.

The concentrations of every species for such a system can be analytically determined. The three equilibrium constants for such a system are shown below:

$$S1 + S2 \rightleftharpoons DH \qquad K_1 = \frac{[DH]}{[S1][S2]} \tag{1}$$

$$S1 \rightleftharpoons H1 \qquad K_2 = \frac{[H1]}{[S1]} \tag{2}$$

$$S2 \rightleftharpoons H2 \qquad K_3 = \frac{[H2]}{[S2]} \tag{3}$$

where S1, S2, H1, H2, and DH are the random coil S1, the random coil S2, the folded strand H1, the folded strand H2, and the double helix DH, respectively. The conservation of S1 and S2 leads to the following equations:

$$C_{S1}^{Total} = S1 + H1 + DH \tag{4}$$

$$C_{S2}^{Total} = S2 + H2 + DH \tag{5}$$

where C_{S1}^{Total} are the total concentrations of S1 and S2. [DH] and [S2] can be expressed as a function of [S1] by substituting the [H1] obtained from Equation 2, in Equation 6, and then substituting the [DH] obtained by Equation 6 in Equation 1.

$$[DH] = C_{S1}^{Total} - [S1] - K_2[S1] \tag{6}$$

$$[S2] = \frac{[DH]}{K_1[S1]} \Rightarrow [S2] = \frac{C_{S1}^{Total} - [S1] - K_2[S1]}{K_1[S1]} \tag{7}$$

Substitution of [H2], [DH], and [S2] from Equations 3, 6 and 7 in Equation 5 leads to an expression of K_1 that can be rearranged as a quadratic equation in [S1]:

$$[S1]^2 (K_1 + K_2 K_1) + [S1](K_1 C_{S2} + K_3 + K_2 K_3 - K_1 C_{S1} + K_2 + 1) - (K_3 + 1) C_{S1} = 0 \tag{8}$$

This equation is simplified by making the following substitutions:

$$a = (K_1 + K_2 K_1) \quad (9)$$

$$b = (K_1 C_{S2}^{Total} + K_3 + K_2 K_3 - K_1 C_{S1}^{Total} + K_2 + 1) \quad (10)$$

$$c = (K_3 + 1) C_{S1}^{Total} \quad (11)$$

The physical solution of the quadratic equation (i.e. positive root) is (Press, 1999):

$$[S1] = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \quad (12)$$

or

$$[S1] = \frac{2c}{-b - \sqrt{b^2 - 4ac}} \quad (13)$$

The second equation has better numerical stability (Press, 1999).

[DH], [S2], [H1], and [H2] can then be calculated using Equations 1-3.

Determination of Net Free Energy

The net free energy of hybridization is calculated as follows:

$$\Delta G^{\circ}_{37net} = -RT \ln K_{net} \quad (14)$$

where

$$K_{net} = \frac{[DH]}{[S1_{single\ stranded}][S2_{single\ stranded}]} \quad (15)$$

where [S1 single stranded] and [S2 single stranded] are the concentrations of S1 and S2 either in the random coil state or the hairpin states, at the temperature of the simulation. Using the conservation of S1 and S2, Equation 14 is rewritten as follows:

$$\Delta G^{\circ}_{T\,net} = -RT \ln \frac{[DH]}{(C_{S1}^{Total} - [DH])(C_{S2}^{Total} - [DH])} \tag{16}$$

Note that $\Delta G^{\circ}_{T\,net}$ has the unusual property that it depends on the total strand concentrations, C_{S1}^{Total} and C_{S2}^{Total}. The net free energy expresses the duplex formation equilibrium free energy corrected for secondary-structure formation in the single strands.

Determination of Net Melting Temperature

If the strands are non self-complementary two cases have to be considered depending on the relative strand concentrations:

1) If S1 is the limiting reagent ($C_{S1}^{Total} < C_{S2}^{Total}$), at T_M:

$$[DH] = \frac{1}{2} C_{S1}^{Total} \tag{17}$$

The concentrations of strands [S1] and [S2] are given by the following relations:

$$C_{S1}^{Total} = \frac{1}{2} C_{S1}^{Total} + K_2[S1] + [S1] \tag{18a}$$

$$[S1] = \frac{C_{S1}^{Total}}{2(K_2 + 1)} \tag{18b}$$

$$C_{S2}^{Total} = \frac{1}{2} C_{S1}^{Total} + K_3[S2] + [S2] \tag{19a}$$

$$[S2] = \frac{C_{S2}^{Total} - \frac{1}{2} C_{S1}^{Total}}{K_3 + 1} \tag{19b}$$

The replacement of [S1] and [S2] in Equation 1 gives:

$$K_1 = \frac{[DH]}{[S1][S2]} = \frac{K_2 + K_2K_3 + K_3 + 1}{C_{S2}^{Total} - \frac{1}{2}C_{S1}^{Total}} \tag{20a}$$

$$0 = \frac{1}{2}C_{S1}^{Total} - C_{S2}^{Total} + \frac{K_2 + K_2K_3 + K_3 + 1}{K_1} \tag{20b}$$

Using the relation $\Delta G°_T = -R\ T\ \ln K$, Equation 20 is arranged as follows:

$$0 = \frac{1}{2}C_{S1}^{Total} - C_{S2}^{Total} + \left(e^{\frac{-\Delta G°_T(2)}{RT}} + e^{\frac{-\Delta G°_T(2)-\Delta G°_T(3)}{RT}} + e^{\frac{-\Delta G°_T(3)}{RT}} + 1 \right) e^{\frac{\Delta G°_T(1)}{RT}} \tag{21}$$

$\Delta G°_T$ can then be decomposed as $\Delta G°_T = \Delta H° - T\ \Delta S°$ (assuming $\Delta C°_{p=0}$) to obtain:

$$0 = \frac{1}{2}C_{S1}^{Total} + C_{S2}^{Total} + e^{\frac{-\Delta H°(2)+\Delta H°(1)}{RT}} e^{\frac{+\Delta S°(2)-\Delta S°(1)}{R}} + e^{\frac{-\Delta H°(2)-\Delta H°(3)+\Delta H°(1)}{RT}} e^{\frac{\Delta S°(2)+\Delta S°(3)-\Delta S°(1)}{R}} + e^{\frac{-\Delta H°(3)+\Delta H°(1)}{RT}} e^{\frac{\Delta S°(3)-\Delta S°(1)}{R}} + e^{\frac{\Delta H°(1)}{RT}} e^{\frac{-\Delta S°(1)}{R}} \tag{22}$$

The above equation can be solved by bisection or other numerical techniques to find T. This solution is the net melting temperature.

2) If S2 is the limiting reagent ($C_{S2}^{Total} < C_{S1}^{Total}$), the following relation can be deduced by a similar approach:

$$0 = \frac{1}{2}C_{S2}^{Total} - C_{S1}^{Total} + \frac{K_3 + K_3K_2 + K_2 + 1}{K_1} \tag{23}$$

Again, application of the bisection method to an equation symmetric to Equation 22 affords the net melting temperature.

If the strand S is self complementary, the reactions are described by the following equilibria:

$$S + S \rightleftharpoons DH \qquad K_1 = \frac{[DH]}{[S]^2} \tag{24}$$

$$S \rightleftharpoons H \qquad K_2 = \frac{[H]}{[S]} \tag{25}$$

$$At\ T_M: [DH] = \frac{1}{4} C_S^{Total} \tag{26}$$

The strand conservation equation is:

$$C_S^{Total} = [S] + [H] + 2[DH] \tag{27}$$

Insertion of [H] and [DH] from Equations 25 and 26 in Equation 27 leads to:

$$C_S^{Total} = 2(K_2 + 1)[S] \Leftrightarrow [S] = \frac{C_S^{Total}}{2(K_2 + 1)} \tag{28}$$

Introduction of [S] in Equation 24 gives:

$$K_1 = \frac{[DH]}{[S]^2} = \frac{(K_2^2 + 1 + 2K_2)}{C_S^{Total}} \Leftrightarrow 0 = K_2^2 + 2K_2 - K_1 C_S^{Total} + 1 \tag{29}$$

Using the relation $\Delta G°_T = -R\ T\ \ln K$, Equation 29 is rearranged as follows:

$$0 = e^{\frac{-2\Delta G°_T(2)}{RT}} + 2e^{\frac{-\Delta G°_T(2)}{RT}} - C_S^{Total} e^{\frac{\Delta G°_T(1)}{RT}} + 1 \tag{30}$$

$\Delta G°_T$ can then be decomposed as $\Delta G°_T = \Delta H° - T\,\Delta S°$ to obtain:

$$0 = e^{\frac{-2\Delta H°(2)}{RT}} e^{\frac{+2\Delta S°(2)}{R}} + 2e^{\frac{-\Delta H°(2)}{RT}} e^{\frac{+\Delta S°(2)}{R}} - C_S^{Total} e^{\frac{-\Delta H°(1)}{RT}} e^{\frac{+\Delta S°(1)}{R}} + 1 \qquad (31)$$

This equation can be solved by bisection to afford the net melting temperature.

An experimentally validated example of the accuracy of the net hybridization thermodynamics is shown in Figure 8 for molecular beacons. At the top of Figure 8 are the predicted thermodynamics for simple duplex formation assuming no competing single strand secondary structure. Using Module 1 of the invention, these results are similar to what would be predicted using other commercial software (such as oligo 6.0), though our thermodynamic database includes the dangling end effects and salt corrections are more accurate than other software. The middle of Figure 8 shows the single strand folding at the molecular beacon as output from DNA-MFOLD. The bottom table of Figure 8 shows the experimentally determined Δ6 (effective) and Tm (effective) published in Bonnet et al. 1999, as well as the effective Tm and Δ6 (effective) predicted with Module 1 using the coupled equilibria calculations. Note the close agreement between experiments and predictions in the bottom table and the disagreement between experiments and the predictions using the naive simple hybridization calculation (top table of Figure 8). Also note the good agreement in the bottom table for the fully matched A-T sequence and mismatch A-A, A-C, and A-6 sequences, thus validating the mismatch parameters.

Further, the net hybridization calculations can be extended to different temperatures as shown in Figure 9, to reveal how the concentrations of all species change with temperature. Given the extinction coefficients and fluorescence quantum yields, the concentration vs. temperature profiles shown in Figure 9 can be used to calculate the fluorescence vs. temperature profile (not shown), thereby allowing the prediction of the temperature which produces the maximum fluorescence signal and minimum background fluorescence signal.

Another manifestation of the concentration calculations is for match vs. mismatch discrimination (Figure 10), whereby the concentrations of all species at all temperatures can be calculated (Figure 11). For the particular case shown, optimal match vs. mismatch discrimination is predicted to occur at 0°C. The concentration calculations can be generalized for cases in which molecules can form many different competing unimolecular, biomolecular, and higher order complexes (Figure 12) using generalized equations such as shown in Figure 13 for the five molecule case, and solved using the algorithm in Figure 7.

Algorithm

The hybridization prediction algorithm of the present invention is based on a nearest-neighbor-model analysis of the sequences. The algorithm accounts for structural motifs including Watson-Crick base pairs (Allawi and SantaLucia, 1997; SantaLucia, 1998; Sugimoto et al., 1995; Xia et al., 1998), single internal mismatches (Allawi and SantaLucia, 1997; Allawi and SantaLucia, 1998; Allawi and SantaLucia, 1998; Allawi and SantaLucia, 1998; Kierzek et al., 1999; Peyret et al., 1999; SantaLucia, 1998), double mismatches (Allawi and SantaLucia, 1997) coaxial-stacking interfaces (disclosed herein) (Walter and Turner, 1994), terminal mismatches (disclosed herein) (Freier et al., 1986) and dangling ends (Bommarito et al., 2000; Freier et al., 1986). Once the motifs are identified and their thermodynamic contributions are added, the sum may be corrected for salt effects (sodium and magnesium) and the net hybridization is calculated when appropriate.

Algorithm Functions

A first or main module of the algorithm calculates the hybridization thermodynamics ($\Delta H°$, $\Delta S°$, $\Delta G°_{37}$, T_M) of a given duplex. Net hybridization accounting for secondary structure in both strands is also calculated.

Parameterization

Parameters are organized in three arrays. The first array contains internal element parameters: Watson-Crick nearest neighbors and single mismatch nearest neighbors. The second array contains terminal element parameters: terminal mismatches and dangling ends. A single parameter is used to account for double mismatches except for tandem G·T mismatches, which are explicitly enumerated (Allawi & SantaLucia, 1997). The third array contains coaxial-stacking parameters (contained herein).

For DNA sequences, the thermodynamic contribution of all Watson-Crick nearest neighbors and single internal mismatches has been systematically studied (Allawi and SantaLucia, 1997; Allawi and SantaLucia, 1998; Allawi and SantaLucia, 1998; Allawi and SantaLucia, 1998; Peyret et al., 1999). A limited number of sequences containing double mismatches has also been studied (Allawi and SantaLucia, 1997). The contributions of dangling ends (Bommarito et al., 2000) have also been systematically analyzed. Salt corrections are available for sodium in the range 0.01 to 1 M (SantaLucia, 1998).

For RNA sequences, the thermodynamic contribution of all Watson-Crick nearest neighbors has been systematically studied (Xia et al., 1998). A limited number of sequences containing single mismatches has also been studied (Kierzek et al., 1999). The contribution of dangling ends and terminal mismatches has also been systematically analyzed (Freier et al., 1986). No salt correction has been developed for RNA and therefore the DNA salt corrections are assumed. These corrections are likely to be deficient in the case of RNA.

For DNA/RNA hybrids, the thermodynamic contribution of all Watson-Crick nearest neighbors has been systematically studied as well as a limited number of sequences containing single mismatches (Sugimoto et al., 1995). As no salt correction has been developed for DNA/RNA hybrids, the DNA corrections are assumed. The applicability of these corrections to DNA/RNA hybrids has not been tested. The parameter arrays are designed to easily accommodate implementation

of new parameters and salt corrections including thermodynamics parameters for modified bases and denaturant effects.

Correction for Hybridization to DNA Microchips

A linear correction of the free energy is implemented in the algorithm of the invention to correct for hybridization to DNA microchips:

$$\Delta G^\circ_{37}(\text{microchip}) = a\Delta G^\circ_{37}(\text{solution}) + b \qquad (32)$$

where a and b are user defined real coefficients. Fotin et al. (Fotin et al., 1998) showed that a linear relationship could be used to relate the free energies obtained for hybridization in solution and on microchip surfaces. However, the relation between thermodynamics measured in solution and thermodynamics measured using microarrays is still unclear and appears to be different depending on the manufacture and type of microarrays.

User Interface: Input and Output

Figure 2a shows the user interface input. The users enter the sequence of each strand, the hybridization conditions (hybridization temperature, strand concentrations, and monovalent cations and concentrations), and thermodynamic corrections for single strand folding. Figure 2b shows the output corresponding to the input in Figure 2a.

The algorithm can be used via the Internet at: http://jsl1.chem.wayne.edu/Hyther/hytherm1main.html. The algorithm may be written in FORTRAN 77 and run on UNIX environment or other languages and environments.

Molecular Beacons

The algorithm may be used to predict the thermodynamics of a set of literature measurements for molecular beacons (Bonnet et al., 1999). Molecular beacons are high specificity probes that are efficient for mutation analysis (Giensendorf et al., 1998) and multiplex detection of single nucleotide variations (Marras et al., 1999). The design and efficiency optimization of these beacons is helped by hybridization thermodynamics prediction. Bonnet et al. studied, the hybridization of the molecular beacon $^{5'}$CGC, TCC, CAA, AAA, AAA, AAA, CCG AGC G$^{3'}$ to a set of four different targets including a perfect match duplex, and three different duplexes containing one mismatch. Free energy and enthalpy for duplex folding may be calculated using the DNA MFOLD program (http://mfold2.wustl.edu/~mfold/dna/form1.cgi). These parameters may then incorporated as secondary structure corrections in Figure 2a.

The software to implement the algorithm may be written in FORTRAN, C^{++}, Visual Basic, HTML, and JAVA script computer languages. Two graphical user interfaces may be provided: Windows application and web browser format. The software may run on IBM/PC, Sun, and Silicon Graphics platforms.

The software may be written in several modules as described below.

A. *Interactive Mode: Command Line Interface in MS-DOS*

MODULE 1 (As Previously Described above)

Function. Module 1 predicts the hybridization thermodynamics of a given duplex (DNA/DNA, RNA/RNA, or DNA/RNA).

Input (Figure 2a)

Input of Sequences

1. Only the following characters are accepted: A, a, C, c, G, g, T, t, U, u, /, *, +. Single blank characters and numbers will be automatically edited, but more than one carriage return is not permitted.

2. If the duplex contains a dangling end on a strand, the sequence of the other strand should contain a * at the corresponding position. (This is very important to include for primer binding to a large target sequence). Note: The top strand must be entered in 5' to 3' orientation, but the bottom strand must be entered in 3' to 5' orientation. Also, a "+" must be added at the end of each sequence. There is a length limit of 1024 characters for sequence entries. In module 1, it is important to be sure that both sequences have the same length.

Example: AAAACCCCTGA+
TTTGGGGAC+

3. Only the bottom strand may contain coaxially stacked nucleotides. A "/" should be inserted at the site of a strand nick (*i.e.* between the coaxially stacked nucleotides). This feature is useful for predicting stacked hybridization stability.

Example: AAAACCCCC+
TTTT/GGGG+

Input of Salt and Strand Concentrations

The monovalent salt should be the sum of all monovalent cation concentrations in a solution in units of molarity. For example, a solution of 100 mM KCl, 50 mM NaCl, 10 mM Na_2PO_4, 0.1 mM Na_2EDTA would account for a total of 0.1702 M monovalent. The thermodynamic predictions are applicable over a salt range of 0.01 to 1 M monovalent cation. The correction applied is from SantaLucia (1998) *Proc. Natl. Acad. Sci.* 95, 1460. The sodium correction applies

for oligonucleotides with fewer than about 30 base pairs. For longer duplexes a polymer correction is required, but this is not currently implemented.

Strand concentrations are entered in units of molarity. The program will accept virtually any physically relevant strand concentration.

Hybridization temperature is in Celsius degrees. The limits are 0 to 100 degrees.

Special corrections for single-stranded secondary structure and for surface corrections for hybridization arrays can be input. The units for input ΔG° are kcal/mol. To determine estimates of single-strand folding energies, see Michael Zuker's RNA or DNA-MFOLD servers (see http://mfold2.wsutl.edu/~mfold/dna/form1.cgi). The current thermodynamic prediction software incorporates the special corrections for single-stranded secondary structure and for surface corrections for hybridization arrays.

For DNA chip arrays, a linear correction can be applied. The user inputs the slope and intercept coefficients. Based on the work of Mirzabekov group, a slope of +1.1 and intercept of +3.2 are appropriate (see Fotin et al. (1998) *Nucleic Acids Res.* 26, 1515-1521).

Output (Figure 2b)

Module 1 outputs the hybridization thermodynamics at 1.0 M NaCl and 37°C (the conditions under which the thermodynamic predictions are most accurate), under the salt temperature conditions specified by the user, and also displays the net hybridization Tm and ΔG° if the user specifies that special corrections are needed (this allows for single-strand secondary structure of both the target and probe DNA to be accounted and for surface effects of chip arrays). Predictions of ΔG°, ΔH°, ΔS°, and Tm are provided.

MODULE 2

Function. Module 2 finds the best primers of given length complementary to a long target nucleic acid. DNA/DNA, RNA/RNA, DNA/RNA hybridization types are accepted. The user selects the number of primers to output, and the program finds the most stable primers and gives their hybridization position and thermodynamics of each primer.

Input (Figure 3a)

Input of Salt and Strand Concentrations

The input of strand and salt concentrations is similar to Module 1.

Input of Sequences

The target sequence is input as in Module 1.

Output (Figure 3b)

Primer Length and Number of Best Primers

Module 2 displays "number of best primers" best primers of length "primer length" in order of decreasing stability.

Output

Module 2 outputs "number of best primers" best primers of length "primer length" in order of decreasing stability along with their hybridization thermodynamics.

MODULE 3

Function. Module 3 walks a given primer along a given target and finds the thermodynamics for the best target/primer complex and for the competitive target/primer complexes: DNA/DNA, RNA/RNA, DNA/RNA, hybridization types are accepted.

Input (Figure 4a)

Input of Sequences

The input is similar to Module 1. The target has to be longer than the primer.

Input of Salt and Strand Concentrations

The input of salt and strand concentrations is similar to Module 1.

Percent Stability p of Alternative Binding Sites Compared to the Most Stable Binding Site

This parameter excludes all competitive sites that are not within the defined percent of the best primer stability. If the best primer stability is -5 kcal/mol and p=10 then any competitive site of energy higher than -5 + (10/100*5) = -4.5 kcal/mol will not be displayed.

Number of Base Pairs Required to Compute the Solution

This parameter excludes all competitive sites that contain less Watson-Crick base pairs than the defined value.

Output (Figure 4b)

Module 3 outputs the best primer binding site and the competitive binding sites that pass the filtering criteria (percent stability p of alternative binding sites compared to the most stable binding site and number of best primers).

MODULE 4

Function. Batch mode calculations (see below).

MODULE 5

Function. Module 5 is a combination of Modules 2 and 3 and finds the n best primers of given length complementary to a given section of a target and display the thermodynamics of the target/primer system(s). Then, each best primer is walked along the whole target to find the competitive hybridization sites. The thermodynamics of the target/primer systems at these alternative sites is then displayed. DNA/DNA, RNA/RNA, DNA/RNA, hybridization types are accepted.

Input (Figure 5a)

Input of Sequences

The target sequence is input as in Module 1.

Input of Salt and Strand Concentrations

The input of salt and strand concentrations is similar to Module 1.

Sequence Section Where to Find the Best Primers

Module 5 finds the best primers in the target region ranking from "position of initial nucleotide" to "position of final nucleotide". Note that Module 5 then looks for competitive sites of each best primers in the whole target.

Percent Stability of Alternative Binding Sites Compared to the Most Stable Binding Site

The function of this parameter is the same as in Module 3. This parameter is input for each best primer corresponding to the "number of best primer" specified.

Number of Base Pairs Required to Compute the Solution

The function of this parameter is the same as in Module 3. This parameter is input for each best primer corresponding to the "number of best primer" specified.

Output (Figure 5b)

Primer Length and Number of Best Primers

Module 5 displays "number of best primers" best primers of length "primer length" by order of decreasing stability.

Output

Module 5 displays "number of best primers" best primers and their competitive sites by order of stability along with their hybridization thermodynamics. The best primer and its ranked competitive hybridization sites are listed first. Then, the second best primer is listed with its competitive hybridization sites.

MODULE 6

Function. Module 6 is similar to Module 3 and walks a given primer along a given target and finds the thermodynamics for the best target/primer complex and for the competitive target/primer complexes: DNA/DNA, RNA/RNA, DNA/RNA, hybridization types are accepted. Then, Module 6 simulates the concentration of every species at every degree from 1 to 100 °C, as illustrated in Figure 6.

Input (Not Shown)

Input of Sequences

The input is similar to Module 1. The target has to be longer than the primer.

Input of Salt and Strand Concentrations

The input of salt and strand concentrations is similar to Module 1.

Percent Stability p of Alternative Binding Sites

Compared to the Most Stable Binding Site

This parameter excludes all competitive sites that are not within the defined percent of the best primer stability. If the best primer stability is -5 kcal/mol and p=10, then any competitive site of energy higher than: -5 + (10/100*5) = -4.5 kcal/mol will not be displayed.

Number of Base Pairs Required to Compute the Solution

This parameter excludes all competitive sites that contain less Watson-Crick base pairs than the defined value.

Correction for Target/Target Interaction, Target folding, Primer/Primer Interaction and Primer Folding

The user is asked if he wants to correct for the interactions above. If the answer is "y", the user is prompted for ΔH°_{37} corresponding to the interaction. Secondary structure thermodynamics can be determined using the Zuker algorithm as discussed in Module 1 section.

Output (Not Shown)

Concentration Output Filename

The results from the concentration simulations (concentration of species at every temperature) are saved in this file.

Output

Module 6 outputs the best primer binding site and the competitive binding sites that pass the filtering criteria (percent stability p of alternative binding sites compared to the most stable binding site and number of best primers). The concentration simulations are saved in a file specified by the user.

MODULE 7

Function. Module 7 is a combination of Modules 2 and 5 and finds the n best primers of given length complementary to a given section of a target and display the thermodynamics of the target/primer system(s). Then, each best primer is walked along the whole target to find the competitive hybridization sites. The thermodynamics of the target/primer systems at these alternative sites is then displayed. DNA/DNA, RNA/RNA, DNA/RNA hybridization types are accepted. Then, Module 7, like Module 6, simulates the concentration of every species at every degree from 1 to 100°C, as illustrated in Figure 6.

Input (Not Shown)

Input of Sequences

The target sequence is input as in Module 1.

Input of Salt and Strand Concentrations

The input of salt and strand concentrations is similar to Module 1.

Sequence Section Where to Find the Best Primers

Module 7 finds best primers in the target region ranking from "position of initial nucleotide" to "position of final nucleotide." Note that Module 7 then looks for competitive sites of each best primers in the whole target.

Percent Stability of Alternative Binding Sites
Compared to the Most Stable Binding Site

The function of this parameter is the same as in Module 3. This parameter is input for each best primer corresponding to the "number of best primer" specified.

Number of Base Pairs Required to Compute the Solution

The function of this parameter is the same as in Module 3. This parameter is input for each best primer corresponding to the "number of best primer" specified.

Correction for Target/Target Interaction, Target folding, Primer/Primer Interaction and Primer Folding

For each best primer, the user is asked if he wants to correct for the interactions above. If the answer is "y", the user is prompted for $\Delta H°$ and $\Delta G°_{37}$ corresponding to the interaction. Secondary structure thermodynamics can be determined using the Zuker algorithm as discussed in Module 1 section.

Concentration Output Filenames

For each best primer, the results from the concentration simulations (concentration of species at every temperature) are saved in this file. The user has to select a different filename for each best primer.

Output (Not Shown)

Output

Primer Length and Number of Best Primers

Module 7 displays "number of best primers" best primers of length "primer length" by order of decreasing stability.

Module 7 displays "number of best primers" best primers and their competitive sites by order of stability along with their hybridization thermodynamics. The best primer and its ranked competitive hybridization sites are listed first. Then, the second best primer is listed with its competitive hybridization sites. For each best primer, a file named by the user contains the concentration simulations.

Module 7 allows the user to design optimal primers for applications where multiple simultaneous hybridization reactions are occurring, including match vs. mismatch hybridization, molecular beacons, DNA oligonucleotide arrays, and multiplex PCR.

One commercially important example for the use of Module 7 for primer design in a complex hybridization solution is Multiplex PCR, as shown in Figure 7. Module 7 allows the user to design optimal primers for Multiplex PCR where multiple primers have equal stabilities in binding to the target DNA. Several primers must be designed to specifically bind to different sites on target DNA at a given temperature with minimal background binding to mismatch sites and with minimal cross-hybridization between pairs of primers.

Module 7 minimizes potential primer dimer formation and mismatch hybridization for all combinations of input primers. Module 7 optimizes primer sequence position, length, and concentration for each primer in relation to all other species in solution and provides a hybridization profile at all temperatures from 0 to 100°C.

Batch Mode

MODULE 4

Function. Module 4 allows any of the previous modules to be run in batch mode using text files to submit the input and having the data output as text files also.

Type of Input Files

There are two types of input files: 1) parameter input file, and 2) sequence input file. Parameter input files describe what modules to run with what hybridization parameters and on how many sequences to run them. Example of parameter input files for each module with comments are given in the "Batch mode parameter files folder." Sequence files contain the sequences that are going to be

hybridized in the conditions described by the parameter input files. Examples of parameter input files for each module with comments are given in the "Batch mode sequence files folder."

Note that a parameter file can successively run different modules on various different sequences.

The user is successively asked for the names of the parameter input file, the sequence input file and the thermodynamic data output file. Note that these files have to be in the directory containing the executable version of the software. Output files will also be created in this same directory. Names of the concentration simulation files are specified in the parameter input files.

Examples of Batch Mode Parameter Files

Comments in parentheses describe the meaning of each entry (note that an actual parameter file must not contain these comments).

```
DUP   (Module 1: Simple duplex calculations)
1     (Number of sequences to apply this parameter file to)
1     (Monovalent cations concentration mol/L))
1     Mg2+ concentration mol/L)
37.0  (Hybridization temperature)
4e-4  (Top strand concentration mol/L)
4e-4  (Bottom strand concentration mol/L)
1     (Correction for microchips: slope)
0     (Correction for microchips: intercept)
0     (Correction for top strand folding: ΔG°37)
0     (Correction for top strand folding: ΔH°37)
0     (Correction for bottom strand folding: ΔG°37)
0     (Correction for bottom strand folding: ΔH°37)
END   (End of file required)
```

NBP (Module 2: N-best primers)
1 (Number of sequences to apply this parameter file to)
1 (Monovalent cations concentration mol/L)
1 (Mg^{2+} concentration mol/L)
37 (Hybridization temperature)
4e-4 (Top strand concentration mol/L)
4e-4 (Bottom strand concentration mol/L)
4 (Primer Length)
3 (Number of best primers)
1 (Correction for microchips: slope)
0 (Correction for microchips: intercept)
END (End of file required)

PWA (Module 3 primer walk match vs. mismatch sites identification)
1 (Number of sequences to apply this parameter file to)
1 (Monovalent cations concentration mol/L)
1 (Mg^{2+} concentration mol/L)
37 (Hybridization temperature)
4e-4 (Top strand concentration mol/L)
4e-4 (Bottom strand concentration mol/L)
90 (Percent window of best primer stability for alternative sites)
2 (Number of WC base pairs required to compute the solution)
cgcg+(Primer sequence, + required)
1 (Correction for microchips: slope)
0 (Correction for microchips: intercept)
END (End of file required)

BPW (Module 5 displays "number of best primers" best primers and their competitive sites by order of stability along with their hybridization thermodynamics)
1 (Number of sequences to apply this parameter file to)
1 (Lower limit of primer search area)
10 (Upper limit of primer search area)

	1	Mg^{2+} concentration mol/L)
	37	(Hybridization temperature)
	4e-4	(Top strand concentration mol/L)
	4e-4	(Bottom strand concentration mol/L)
5	4	(Primer length)
	1	(Number of best primers)
	1	(Correction for microchips: slope)
	0	(Correction for microchips: intercept)
	800	(Percent window of best primer stability for alternative sites)
10	2	(Number of WC base pairs required to compute the solution)
	END	(End of file required)

	PWC	(Module 6 primer walk with concentration calculations)
	1	(Number of sequences to apply this parameter file to)
	1	(Monovalent cations concentration mol/L)
15	1	Mg^{2+} concentration mol/L)
	37	(Hybridization temperature)
	4e-4	(Top strand concentration mol/L)
	4e-4	(Bottom strand concentration mol/L)
	90	(Percent window of best primer stability for alternative sites)
20	2	(Number of WC base pairs required to compute the solution)
	cgcg+	(Primer sequence, + required)
	1	(Correction for microchips: slope)
	0	(Correction for microchips: intercept)
	0	(Correction for target folding: $\Delta G°_{37}$)
25	0	(Correction for target folding: $\Delta H°$)
	0	(Correction for target/target interaction: $\Delta G°_{37}$)
	0	(Correction for target/target interaction: $\Delta H°$)
	0	(Correction for primer folding: $\Delta G°_{37}$)
	0	(Correction for primer folding: $\Delta H°$)
30	0	(Correction for primer/primer interaction: $\Delta G°_{37}$)
	0	(Correction for primer/primer interaction: $\Delta H°$)
	outconc	(Concentration output file name)

	END	(End of file required)
	BWC	(Module 7, N-best primers, primer walk, and concentration calculations)
	1	(Number of sequences to apply this parameter file to)
5	1	(Lower limit of primer search area)
	10	(Upper limit of primer search area)
	1	(Monovalent cations concentration mol/L)
	0	Mg^{2+} concentration mol/L)
	37	(Hybridization temperature)
10	4e-4	(Top strand concentration mol/L)
	4e-4	(Bottom strand concentration mol/L)
	4	(Primer length)
	1	(Number of best primers)
	1	(Correction for microchips: slope)
15	0	(Correction for microchips: intercept)
	800	(Percent window of best primer stability for alternative sites)
	2	(Number of WC base pairs required to compute the solution)
	0	(Correction for target folding: $\Delta G°_{37}$)
	0	(Correction for target folding: $\Delta H°$)
20	0	(Correction for target/target interaction: $\Delta G°_{37}$)
	0	(Correction for target/target interaction: $\Delta H°$)
	0	(Correction for primer folding: $\Delta G°_{37}$)
	0	(Correction for primer folding: $\Delta H°$)
	0	(Correction for primer/primer interaction: $\Delta G°_{37}$)
25	0	(Correction for primer/primer interaction: $\Delta H°$)
	outconc	(Concentration output file name)
	END	(End of file required)
	PPW	(Module 8: walk a primer along itself to find interaction sites. PWA with probe = primer)
30	1	(Number of sequences to apply this parameter file to)
	1	(Monovalent cations concentration mol/L)

	1	Mg^{2+} concentration mol/L)
	37	(Hybridization temperature)
	4e-4	(Primer concentration mol/L)
	900	(Percent window of best primer stability for alternative sites)
5	2	(Number of WC base pairs required to compute the solution)
	END	(End of file required)

Examples of Batch Mode Sequence Files

For Module 1: Dup

	1	(Sequence number)
10	agcgca+	(Top strand sequence)
	tcgcgt+	(Bottom strand sequence)

For Module 2: NBP

1	(Sequence number)
agcgca+	(Target sequence)

15 For Module 3:

1	(Sequence number)
cgcctgcggccc+	(Target sequence)

For Module 5: bpw

	1	(Sequence number)
20	cgcctgcgccc+	(Target sequence)

For Module 6: pwc

1	(Sequence number)
agcgca+	(Target sequence)

For Module 7: bwc

25	1	(Sequence number)
	agcgca+	(Target sequence)

For Module 8: ppw

1 (Sequence number)

agcgca+ (Primer sequence)

Example of Batch Mode Parameter and Sequence Files to Run Different Modules Successively

Parameter File:

	DUP	(Executes Module 1)
	2	(Apply to Module 1 to 2 sequence sets)
	0.05	
	1.5e-3	
	37.0	
	1e-6	
	2e-7	
	1	
	0.	
	-2.12	
	-37.3	
	0	
	0	
	PWC	(Executes Module 6)
	1	(Apply to Module 6 to 1 sequence set)
	0.16	
	0.0025	
	37	
	10e-9	
	1e-9	
	800	
	8	
	TCGAACGTAC+	
	1	

```
0
0
0
0
0
0
0
0
0
0
outwash
DUP                (Executes Module 1)
4                  (Apply to Module 1 to 4 sequence sets)
1
0
37.0
1e-6
1e-6
1
0
0
0
0
0
END
```

Other modules can be similarly appended.

Sequence File

```
1                                          (input for Module 1)
ttgcctaggggaccaggtccaact+
aacggatcccctggtccaggttga+
2                                          (input for Module 1)
```

ttgcctaggggaccaggtccaact+
aacggatcccctggtccaggttga+
3 (input for Module 6)
CAGCTTGCATGAAAAGCTTGCGTGT+
4 (input for Module 1)
AAAAAA+
TTTTTT+
5 (input for Module 1)
acgcgc+
tgcgcg+
6 (input for Module 1)
gggaaagggg+
cctttccc+
7 (input for Module 1)
tttaaattt+
aaatttaaa+
8 (input for Module 1)
cgcgtgagggcc+
gcgctctccccgg+

Parameterization of the Algorithm of the Invention

Caution: RNA/RNA and DNA/RNA duplexes contain motifs for which no literature data are available. In these cases, DNA/DNA parameters are assumed. Therefore, predictions might be inaccurate. Users are encouraged to use this program with caution and discernment.

No data are available for the following motifs:

RNA/RNA single mismatches

RNA/DNA single mismatches
dangling ends
terminal mismatches
Single mismatches

Double mismatch parameters are estimated for all types of duplexes (DNA/DNA, RNA/RNA, DNA/RNA).

DNA THERMODYNAMIC PARAMETERS		
Watson-Crick nearest-neighbors SantaLucia, Allawi, and Seneviratne (1996) *Biochemistry* 35, 3555; Allawi and SantaLucia (1997) *Biochemistry* 36, 10581	12 parameters	108 sequences
Single mismatch nearest-neighbors Allawi and SantaLucia (1997) *Biochemistry* 36, 100581; Allawi and SantaLucia (1997) *Nucleic Acids Res.*, 26, 2694; Peyret et al., (1999) *Biochemistry* 38, 3468	44 parameters Allawi and SantaLucia (1998) *Biochemistry* 37, 2170 Allawi and SantaLucia (1998) *Biochemistry* 37, 9435	180 sequences
Terminal Mismatch nearest-neighbors (Appendix)	48 parameters	48 sequences
Dangling end nearest-neighbors S. Bommarito, Peyret, SantaLucia (2000) *Nucleic Acids Res.* 28, 1929-1934.	16 parameters	16 sequences
Na^+ dependence law SantaLucia (1998) *Proc. Natl. Acad. Sci.* USA 95, 1460-1465	1 parameter	86 sequences

RNA THERMODYNAMIC PARAMETERS	
Watson-Crick nearest-neighbors Xia et al. (1998) *Biochemistry* 37, 14719	12 parameters
Single Mismatches Kierzek, Burkard, and Turner (1999) *Biochemistry* 38, 14214-14223	44 parameters
Terminal mismatch nearest-neighbors Freier et al. (1986) *Proc. Natl. Acad. Sci.* USA 83, 9373	48 parameters
Dangling end nearest-neighbors Freier et al. (1986) *Proc. Natl. Acad. Sci.* USA 83, 9373	32 parameters
Coaxial stacking nearest-neighbors Walter and Turner (1994) *Biochemistry* 33, 12715	16 parameters
Loop parameters Matthews et al. (1999) *J. Mol. Biol.* 288, 911-940	

HYBRID DNA/RNA THERMODYNAMIC PARAMETERS	
Watson-Crick nearest-neighbors Sugimoto et al. (1995) *Biochemistry* 34, 11211	17 parameters
rU•dG and rG•dT mismatches Sugimoto et al. (1997) *Nucleic Acids Symp. Ser.* 37, 199	

DNA LOOP THERMODYNAMIC PARAMETERS
• Hairpins Hilbers et al. (1985) *Biochie* 67, 685-695 Blommers et al. (1989) *Biochemistry* 28, 7491-7498 Antao et al. (1991) *Nucleic Acids Res. 19, 5901-5905* Antao et al. (1991) *Nucleic Acids Res. 20, 819-824* Senior et al. (1988) *Proc. Natl. Acad. Sci.* USA 85, 6242-6246
• Bulges LeBlanc and Morden (1991) *Biochemistry* 30, 4042-4047 Zieba et al. (1991) *Biochemistry* 30, 8018-8026 Ke et al. (1995) *Biochemistry* 34, 4593-4600 Turner, D.H. (1992) *Curr. Opin. Struc. Biol.* 2, 334-337
• Multibranched Loops Kadrmas et al. (1995) *Nucleic Acids Res.* 23, 2122 Lilley and Hallam (1984) *J. Mol. Biol.* 180, 179-200 Lu et al. (1991) *J. Mol. Biol.* 223, 781-789 Ladbury et al. (1994) *Biochemistry* 33, 6828-6833 Leontis et al. (1991) *Nucleic Acids Res.* 19, 759-766
The parameters for multibranched loops are from a best fit analysis of secondary structure predictions vs. experiments as done by Jaeger et al. for RNA (Jaeger et al. (1989) PNAS 86, 7706-7710). The current parameters for multibranched loops neglect the sequence and complicated length dependence described by Leontis and coworkers, but approximate 4-way junctions fairly well. Implementation of more complicated rules will require modification of the MFOLD algorithm.

While the best mode for carrying out the invention has been described in detail, those familiar with the art to which this invention relates will recognize various alternative designs and embodiments for practicing the invention as defined by the following claims.

APPENDIX

Table 1: Thermodynamic Parameters for Duplex Formation in 1M NaCl[a].

	ΔH^{o} [b] (kcal / mol)	ΔS^{o} [b] (cal / mol K)	ΔG^{o}_{37} [b] (kcal / mol)	T_M [c] (°C)
A T G A G C T C A **A** **A** A C T C G A G T **A**	-56.7 ± 2.5	-155.8 ± 3.7	-9.07 ± 0.12	52.3
A A G A G C T C T **A** **A** T C T C G A G A **A**	-55.1 ± 1.4	-149.2 ± 3.8	-8.91 ± 0.12	56.0
A G T A G C T A C **A** **A** C A T C A A T G **A**	-60.3 ± 1.8	-165.8 ± 5.1	-9.03 ± 0.12	54.5
A C G A T A T C G **A** **A** G C T A T A G C **A**	-67.8 ± 1.3	-192.0 ± 3.5	-8.87 ± 0.06	49.4
C T G A G C T C A **C** **C** A C T C G A G T **C**	-50.6 ± 1.2	-136.9 ± 2.8	-8.15 ± 0.11	52.9
C A G A G C T C T **C** **C** T C T C G A G A **C**	-51.4 ± 1.3	-137.5 ± 3.2	-8.34 ± 0.14	56.9
C G T A G C T A C **C** **C** C A T C G A T G **C**	-55.8 ± 1.8	-154.6 ± 4.8	-8.04 ± 0.16	49.7
C C G A T A T C G **C** **C** G C T A T A G C **C**	-59.8 ± 1.2	-166.9 ± 3.0	-8.12 ± 0.07	49.7
G T G A G C T C A **G** **G** A C T C G A G T **G**	-52.6 ± 1.3	-140.7 ± 3.3	-8.57 ± 0.13	57.7
G A G A G C T C T **G** **G** T C T C G A G A **G**	-52.3 ± 1.4	-142.5 ± 4.1	-8.34 ± 0.08	52.3
G G T A G C T A C **G** **G** C A T C G A T G **G**	-59.2 ± 1.0	-164.2 ± 2.8	-8.68 ± 0.07	51.2
G C G A T A T C G **G** **G** G C T A T A G C **G**	-65.7 ± 0.9	-183.6 ± 2.2	-8.81 ± 0.07	52.2

Table 1: Continued[a].

	ΔH° [b] (kcal / mol)	ΔS° [b] (cal / mol K)	ΔG°_{37} [b] (kcal / mol)	T_M [c] (°C)
T T G A G C T C A **T** **T** A C T C G A G T **T**	-55.4 ± 1.1	-149.4 ± 3.0	-8.62 ± 0.10	56.9
T A G A G C T C T **T** **T** T C T C G A G A **T**	-56.5 ± 1.3	-154.3 ± 3.7	-8.72 ± 0.09	54.3
T G T A G C T A C **T** **T** C A T C G A T G **T**	-63.8 ± 0.8	-178.0 ± 2.1	-8.75 ± 0.06	52.1
T C G A T A T C G **T** **T** G C T A T A G C **T**	-66.8 ± 0.6	-187.9 ± 1.6	-8.60 ± 0.04	50.5
C T G A G C T C A **A** **A** A C T C G A G T **C**	-53.6 ± 1.3	-145.8 ± 4.0	-8.42 ± 0.06	53.6
A T G A G C T C A **C** **C** A C T C G A G T **A**	-54.0 ± 1.3	-144.4 ± 3.2	-8.92 ± 0.15	58.8
C G T A G C T A C **A** **A** C A T C G A T G **C**	-56.8 ± 1.4	-155.6 ± 3.5	-8.53 ± 0.13	53.7
A G T A G C T A C **C** **C** C A T C G A T G **A**	-57.1 ± 1.2	-156.0 ± 3.0	-8.71 ± 0.16	54.4
C C G A T A T C G **A** **A** G C T A T A G C **C**	-61.8 ± 0.5	-172.7 ± 1.4	-8.30 ± 0.03	50.6
A C G A T A T C G **C** **C** G C T A T A G C **A**	-58.3 ± 1.7	-158.5 ± 4.2	-8.91 ± 0.17	56.8
C A G A G C T C T **A** **A** T C T C G A G A **C**	-54.6 ± 0.6	-147.9 ± 1.4	-8.66 ± 0.07	55.4

Table 1: Continued[a].

	ΔH° [b] (kcal / mol)	ΔS° [b] (cal / mol K)	ΔG°_{37} [b] (kcal / mol)	T_M [c] (°C)
A A G A G C T C T **C** **C** T C T C G A G A **A**	-55.5 ± 1.4	-150.4 ± 4.4	-8.85 ± 0.08	55.8
T T G A G C T C A **C** **C** A C T C G A G T **T**	-52.2 ± 1.0	-140.8 ± 2.4	-8.38 ± 0.11	54.9
C T G A G C T C A **T** **T** A C T C G A G T **C**	-55.1 ± 0.8	-150.5 ± 1.9	-8.42 ± 0.08	53.2
T G T A G C T A C **C** **C** C A T C G A T G **T**	-58.0 ± 1.4	-159.6 ± 3.7	-8.39 ± 0.10	52.9
C G T A G C T A C **T** **T** C A T C G A T G **C**	-59.4 ± 0.9	-164.6 ± 2.1	-8.21 ± 0.06	51.7
T C G A T A T C G **C** **C** G C T A T A G C **T**	-61.7 ± 1.1	-170.6 ± 3.2	-8.33 ± 0.12	53.7
C C G A T A T C G **T** **T** G C T A T A G C **C**	-57.9 ± 1.1	-159.5 ± 2.7	-8.11 ± 0.12	52.7
T A G A G C T C T **C** **C** T C T C G A G A **T**	-55.0 ± 1.4	-148.0 ± 3.5	-8.80 ± 0.12	57.5
C A G A G C T C T **T** **T** T C T C G A G A **C**	-51.5 ± 1.1	-137.9 ± 2.6	-8.44 ± 0.15	56.4
G T G A G C T C A **A** **A** A C G C G A G T **G**	-54.2 ± 1.4	-146.1 ± 3.3	-8.77 ± 0.16	56.8
A T G A G C T C A **G** **G** A C T C G A G T **A**	-55.4 ± 1.2	-148.6 ± 3.1	-9.03 ± 0.14	59.1

Table 1: Continued[a].

	ΔH° [b] (kcal / mol)	ΔS° [b] (cal / mol K)	ΔG°_{37} [b] (kcal / mol)	T_M [c] (°C)
G G T A G C T A C **A** **A** C A T C G A T G **G**	-59.4 ± 1.6	-163.3 ± 3.9	-8.76 ± 0.15	53.8
A G T A G C T A C **G** **G** C A T C G A T G **A**	-63.7 ± 1.6	-174.9 ± 4.1	-9.46 ± 0.18	56.4
G C G A T A T C G **A** **A** G C T A T A G C **G**	-60.4 ± 0.8	-166.6 ± 2.1	-8.50 ± 0.09	53.5
A C G A T A T C G **G** **G** G C T A T A G C **A**	-61.1 ± 1.3	-167.5 ± 3.4	-9.04 ± 0.14	55.8
G A G A G C T C T **A** **A** T C T C G A G A **G**	-54.0 ± 1.1	-144.9 ± 2.7	-8.82 ± 0.14	57.9
A A G A G C T C T **G** **G** T C T C G A G A **A**	-54.8 ± 1.7	-148.0 ± 5.0	-8.90 ± 0.09	56.3
T T G A G C T C A **G** **G** A C T C G A G T **T**	-56.8 ± 0.7	-155.5 ± 1.7	-8.64 ± 0.06	53.5
G T G A G C T C A **T** **T** A C T C G A G T **G**	-57.4 ± 0.7	-156.7 ± 2.0	-8.80 ± 0.05	54.7
T G T A G C T A C **G** **G** C A T C G A T G **T**	-59.2 ± 1.3	-161.3 ± 3.6	-8.93 ± 0.08	56.2
G G T A G C T A C **T** **T** C A T C G A T G **G**	-64.8 ± 2.6	-182.3 ± 7.2	-8.63 ± 0.17	50.0
T C G A T A T C G **G** **G** C C T A T A G C **T**	-63.3 ± 0.6	-176.9 ± 1.6	-8.42 ± 0.05	51.2
G C G A T A T C G **T** **T** G C T A T A G C **G**	-63.7 ± 0.9	-177.4 ± 2.2	-8.73 ± 0.12	52.6

Table 1: Continued[a].

	ΔH^{o} [b] (kcal / mol)	ΔS^{o} [b] (cal / mol K)	ΔG^{o}_{37} [b] (kcal / mol)	T_M [c] (°C)
T A G A G C T C T **G** **G** T C T C G A G A **T**	-57.8 ± 0.6	-157.4 ± 1.7	-8.95 ± 0.03	55.6
G A G A G C T C T **T** **T** T C T C G A G A **G**	-57.3 ± 1.2	-155.5 ± 3.3	-8.95 ± 0.10	56.4
Core sequences				
C G A T A T C G [d] G C T A T A G C	-51.9 ± 0.6	-145.3 ± 1.4	-6.89 ± 0.09	44.1
G T A G C T A C [d] C A T C G A T G	-51.6 ± 0.6	-143.7 ± 1.3	-7.01 ± 0.08	45.6
A G A G C T C T T C T C G A G A	-50.0 ± 0.7	-136.5 ± 1.7	-7.76 ± 0.06	50.2
T G A G C T C A A C T C G A G T	-50.5 ± 0.5	-137.7 ± 1.3	-7.73 ± 0.04	50.4

[a] The top strand of each duplex is represented in the 5' to 3' orientation and the bottom strand is shown in the 3' to 5' direction. Terminal mismatch nearest neighbors are represented in bold. Mismatches are underlined. [b] ΔH^{o}, ΔS^{o}, and ΔG^{o}_{37} are the error-weighted averages of the $1/T_M$ vs. ln C_T plot and curve fit methods in Table S1. Errors reflect the precision of the data (see text). [c] T_M calculated using 10^{-4} M total strand concentration. [d] Data from reference (19).

Table 2: Nearest-neighbor thermodynamic paramters of like-with-like base terminal mismatches in 1 M NaCl

Dimer Sequence[a]	ΔH° [b] (kcal/mol)	ΔS° [b] (e.u)	ΔG°_{37} [b] (kcal/mol)
	Terminal A•A Mismatches		
A<u>A</u>/T<u>A</u>	-3.1 ± 1.3	-7.8 ± 2.0	-0.67 ± 0.06
T<u>A</u>/A<u>A</u>	-2.5 ± 0.8	-6.3 ± 2.1	-0.58 ± 0.07
C<u>A</u>/G<u>A</u>	-4.3 ± 1.0	-10.7 ± 2.6	-1.01 ± 0.07
G<u>A</u>/C<u>A</u>	-8.0 ± 0.7	-22.5 ± 1.9	-0.99 ± 0.06
	Terminal C•C Mismatches		
A<u>C</u>/T<u>C</u>	-0.1 ± 0.6	0.5 ± 1.5	-0.21 ± 0.06
T<u>C</u>/A<u>C</u>	-0.7 ± 0.7	-1.3 ± 1.8	-0.29 ± 0.07
C<u>C</u>/G<u>C</u>	-2.1 ± 0.9	-5.1 ± 2.5	-0.52 ± 0.09
G<u>C</u>/C<u>C</u>	-3.9 ± 0.7	-10.6 ± 1.7	-0.62 ± 0.06
	Terminal G•G Mismatches		
A<u>G</u>/T<u>G</u>	-1.1 ± 0.7	-2.1 ± 1.8	-0.42 ± 0.07
T<u>G</u>/A<u>G</u>	-1.1 ± 0.8	-2.7 ± 2.2	-0.29 ± 0.05
C<u>G</u>/G<u>G</u>	-3.8 ± 0.6	-9.5 ± 1.5	-0.83 ± 0.05
G<u>G</u>/C<u>G</u>	-0.7 ± 0.5	-19.2 ± 1.3	-0.96 ± 0.06
	Terminal T•T Mismatches		
A<u>T</u>/T<u>T</u>	-2.4 ± 0.6	-6.5 ± 1.6	-0.45 ± 0.05
T<u>T</u>/A<u>T</u>	-3.2 ± 0.7	-8.9 ± 2.1	-0.48 ± 0.05
C<u>T</u>/G<u>T</u>	-6.1 ± 0.5	-16.9 ± 1.2	-0.87 ± 0.05
G<u>T</u>/C<u>T</u>	-7.4 ± 0.4	-21.2 ± 1.1	-0.86 ± 0.05

[a] Thermodynamic parameters and their corresponding errors are calculated from Table 1 using equations 4 and 5.

[b] Dimers are given in antiparallel orientation (e.g. A<u>C</u>/T<u>A</u> equals 5'-AC-3' paired with 3'-TA-5'). Mismatches are underlined.

Table 3: Nearest-neighbor thermodynamic parameters for mixed-base terminal mismatches in 1 M NaCl

Dimer sequence[a]	ΔH° [b] (kcal/mol)	ΔS° [b] (e.u)	ΔG°_{37} [b] (kcal/mol)
	Terminal A•C Mismatches		
AA/TC	-1.6 ± 0.7	-4.0 ± 2.1	-0.35 ± 0.04
AC/TA	-1.8 ± 0.7	-3.8 ± 1.7	-0.59 ± 0.08
CA/GC	-2.6 ± 0.8	-5.9 ± 1.8	-0.76 ± 0.07
CC/GA	-2.7 ± 0.7	-6.0 ± 1.6	-0.85 ± 0.09
GA/CC	-5.0 ± 0.4	-13.8 ± 1.0	-0.71 ± 0.05
GC/CA	-3.2 ± 0.9	-7.1 ± 2.2	-1.01 ± 0.10
TA/AC	-2.3 ± 0.5	-5.9 ± 1.1	-0.45 ± 0.05
TC/AA	-2.7 ± 0.8	-7.0 ± 2.4	-0.55 ± 0.05
	Terminal C•T Mismatches		
AC/TT	-0.9 ± 0.5	-1.7 ± 1.4	-0.33 ± 0.06
AT/TC	-2.3 ± 0.5	-6.3 ± 1.2	-0.35 ± 0.05
CC/GT	-3.2 ± 0.8	-8.0 ± 2.0	-0.69 ± 0.07
CT/GC	-3.9 ± 0.6	-10.6 ± 1.2	-0.60 ± 0.05
GC/CT	-4.9 ± 0.6	-13.5 ± 1.7	-0.72 ± 0.08
GT/CC	-3.0 ± 0.6	-7.8 ± 1.5	-0.61 ± 0.08
TC/AT	-2.5 ± 0.8	-6.3 ± 2.0	-0.52 ± 0.07
TT/AC	-0.7 ± 0.6	-1.2 ± 1.6	-0.34 ± 0.08
	Terminal G•A Mismatches		
AA/TG	-1.9 ± 0.7	-4.4 ± 1.8	-0.52 ± 0.08
AG/TA	-2.5 ± 0.7	-5.9 ± 1.7	-0.65 ± 0.07
CA/GG	-3.9 ± 0.8	-9.6 ± 2.1	-0.88 ± 0.09
CG/GA	-6.0 ± 0.9	-15.5 ± 2.1	-1.23 ± 0.1
GA/CG	-4.3 ± 0.5	-11.1 ± 1.3	-0.80 ± 0.06
GG/CA	-4.6 ± 0.7	-11.4 ± 1.8	-1.08 ± 0.09
TA/AG	-2.0 ± 0.7	-4.7 ± 1.6	-0.53 ± 0.07
TG/AA	-2.4 ± 0.9	-5.8 ± 2.7	-0.57 ± 0.05

Table 3: Continued

Dimer sequence[a]	$\Delta H°$ [b] (kcal/mol)	$\Delta S°$ [b] (e.u)	$\Delta G°_{37}$ [b] (kcal/mol)
	Terminal G•T Mismatches		
AG/TT	-3.2 ± 0.4	-8.7 ± 1.1	-0.45 ± 0.04
AT/TG	-3.5 ± 0.4	-9.4 ± 1.2	-0.54 ± 0.03
CG/GT	-3.8 ± 0.7	-9.0 ± 1.9	-0.96 ± 0.06
CT/GG	-6.6 ± 1.3	-18.7 ± 3.6	-0.81 ± 0.09
GG/CT	-5.7 ± 0.4	-15.9 ± 1.0	-0.76 ± 0.05
GT/CG	-5.9 ± 0.5	-16.1 ± 1.3	-0.92 ± 0.07
TG/AT	-3.9 ± 0.5	-10.5 ± 1.2	-0.59 ± 0.03
TT/AG	-3.6 ± 0.7	-9.8 ± 1.9	-0.59 ± 0.06

[a] Thermodynamic parameters and their corresponding errors are calculated from Table 1 using equations 4 and 5.

[b] Dimers are given in antiparallel orientation (e.g. AC/TA equals 5'-AC-3' paired with 3'-TA-5'). Mismatches are underlined.

Table S1: Thermodynamic Parameters for Duplex Formation in 1M NaCl[a].

	ΔH° (kcal / mol)	ΔS° (cal / mol K)	ΔG°_{37} (kcal / mol)	T_M[b] (°C)
A T G A G C T C A **A**	[c] -56.0 ± 3.2	-151.4 ± 10.1	-9.07 ± 0.13	57.0
A A C T C G A G T **A**	[d] -57.6 ± 4.0	-156.5 ± 4.0	-9.11 ± 0.46	56.7
A A G A G C T C T **A**	[c] -54.6 ± 1.8	-146.9 ± 5.3	-8.91 ± 0.12	57.7
A T C T C G A G A **A**	[d] -56.0 ± 2.3	-151.6 ± 5.4	-8.97 ± 0.64	56.4
A G T A G C T A C **A**	[c] -59.9 ± 2.2	-164.0 ± 6.6	-9.03 ± 0.12	55.4
A C A T C A A T G **A**	[d] -61.3 ± 3.3	-168.3 ± 7.9	-9.09 ± 0.83	55.3
A C G A T A T C G **A**	[c] -60.5 ± 2.4	-166.5 ± 7.7	-8.86 ± 0.06	54.3
A G C T A T A G C **A**	[d] -70.8 ± 1.5	-198.4 ± 3.9	-9.28 ± 0.34	53.6
C T G A G C T C A **C**	[c] -52.3 ± 4.0	-142.4 ± 12.7	-8.16 ± 0.11	52.4
C A C T C G A G T **C**	[d] -50.4 ± 1.2	-136.6 ± 2.8	-8.06 ± 0.36	52.4
C A G A G C T C T **C**	[c] -55.6 ± 2.5	-152.1 ± 7.7	-8.36 ± 0.14	52.8
C T C T C G A G A **C**	[d] -49.9 ± 1.5	-134.5 ± 3.5	-8.14 ± 0.45	53.1
C G T A G C T A C **C**	[c] -55.1 ± 2.4	-151.8 ± 7.2	-8.04 ± 0.16	50.9
C C A T C G A T G **C**	[d] -56.8 ± 2.7	-156.9 ± 6.6	-8.08 ± 0.70	50.7
C C G A T A T C G **C**	[c] -57.3 ± 2.6	-158.4 ± 8.1	-8.12 ± 0.07	50.8
C G C T A T A G C **C**	[d] -60.5 ± 1.3	-168.3 ± 3.3	-8.21 ± 0.31	50.9
G T G A G C T C A **G**	[c] -55.2 ± 2.1	-150.4 ± 6.4	-8.58 ± 0.14	54.2
G A C T C G A G T **G**	[d] -50.9 ± 1.7	-137.1 ± 3.9	-8.38 ± 0.48	54.4
G A G A G C T C T **G**	[c] -56.8 ± 2.0	-155.1 ± 5.6	-8.72 ± 0.26	54.6
G T C T C G A G A **G**	[d] -48.0 ± 1.9	-128.1 ± 5.9	-8.30 ± 0.09	54.9

Table S1: Continued.

	ΔH° (kcal / mol)	ΔS° (cal / mol K)	ΔG°_{37} (kcal / mol)	T_M[b] (°C)
G G T A G C T A C **G**	[c] -56.9 ± 1.5	-155.4 ± 4.8	-8.67 ± 0.07	54.2
G C A T C G A T G **G**	[d] -61.1 ± 1.4	-168.5 ± 3.4	-8.83 ± 0.35	53.9
G C G A T A T C G **G**	[c] -62.4 ± 2.4	-172.9 ± 7.4	-8.79 ± 0.08	53.3
G G C T A T A G C **G**	[d] -66.2 ± 0.9	-184.7 ± 2.3	-8.93 ± 0.21	53.0
T T G A G C T C A **T**	[c] -56.8 ± 1.5	-155.1 ± 4.7	-8.63 ± 0.10	54.2
T A C T C G A G T **T**	[d] -53.7 ± 1.7	-145.6 ± 3.9	-8.52 ± 0.46	54.4
T A G A G C T C T **T**	[c] -58.7 ± 1.6	-160.3 ± 4.6	-8.96 ± 0.19	55.4
T T C T C G A G A **T**	[d] -52.9 ± 2.1	-142.5 ± 6.4	-8.66 ± 0.10	55.6
T G T A G C T A C **T**	[c] -63.2 ± 1.1	-175.5 ± 3.4	-8.75 ± 0.06	52.9
T C A T C G A T G **T**	[d] -64.4 ± 1.1	-179.4 ± 2.6	-8.80 ± 0.25	52.8
T C G A T A T C G **T**	[c] -64.6 ± 1.3	-180.4 ± 4.0	-8.59 ± 0.05	51.7
T G C T A T A G C **T**	[d] -67.4 ± 0.7	-189.4 ± 1.8	-8.69 ± 0.16	51.5
C T G A G C T C A **A**	[c] -56.0 ± 2.1	-153.1 ± 6.2	-8.58 ± 0.15	53.9
A A C T C G A G T **C**	[d] -52.0 ± 1.7	-140.6 ± 5.3	-8.39 ± 0.07	54.1
A T G A G C T C A **C**	[c] -57.0 ± 2.4	-154.9 ± 7.1	-8.94 ± 0.16	55.8
C A C T C G A G T **A**	[d] -52.7 ± 1.6	-141.7 ± 3.6	-8.73 ± 0.46	56.1
C G T A G C T A C **A**	[c] -57.8 ± 2.9	-158.8 ± 8.9	-8.54 ± 0.13	53.2
A C A T C G A T G **C**	[d] -56.5 ± 1.6	-155.0 ± 3.8	-8.46 ± 0.42	53.0
A G T A G C T A C **C**	[c] -59.4 ± 4.0	-163.2 ± 12.2	-8.74 ± 0.18	53.9
C C A T C G A T G **A**	[d] -56.8 ± 1.3	-155.5 ± 3.1	-8.60 ± 0.35	53.8
C C G A T A T C G **A**	[c] -61.6 ± 1.1	-171.8 ± 3.5	-8.30 ± 0.03	50.8
A G C T A T A G C **C**	[d] -61.9 ± 0.6	-172.9 ± 1.5	-8.30 ± 0.14	50.7

Table S1: Continued.

	ΔH° (kcal / mol)	ΔS° (cal / mol K)	ΔG°_{37} (kcal / mol)	T_M [b] (°C)
A C G A T A T C G **C**	[c] -61.8 ± 3.2	-170.3 ± 9.6	-8.94 ± 0.19	54.3
C G C T A T A G C **A**	[d] -57.0 ± 1.9	-155.8 ± 4.6	-8.71 ± 0.52	54.4
C A G A G C T C T **A**	[c] -55.7 ± 1.1	-151.6 ± 3.5	-8.67 ± 0.07	54.6
A T C T C G A G A **C**	[d] -54.2 ± 0.7	-147.2 ± 1.6	-8.59 ± 0.18	54.6
A A G A G C T C T **C**	[c] -58.6 ± 4.1	-159.8 ± 12.4	-9.07 ± 0.26	56.1
C T C T C G A G A **A**	[d] -55.1 ± 1.5	-149.1 ± 4.7	-8.83 ± 0.08	55.8
T T G A G C T C A **C**	[c] -54.1 ± 1.9	-147.4 ± 5.8	-8.40 ± 0.12	53.4
C A C T C G A G T **T**	[d] -51.5 ± 1.1	-139.4 ± 2.6	-8.27 ± 0.32	53.5
C T G A G C T C A **T**	[c] -57.3 ± 4.5	-157.5 ± 14.3	-8.43 ± 0.09	52.7
T A C T C G A G T **C**	[d] -55.0 ± 0.8	-150.3 ± 1.9	-8.37 ± 0.22	53.0
T G T A G C T A C **C**	[c] -58.7 ± 2.1	-162.3 ± 6.5	-8.39 ± 0.10	52.0
C C A T C G A T G **T**	[d] -57.4 ± 1.9	-158.2 ± 4.6	-8.32 ± 0.48	52.0
C G T A G C T A C **T**	[c] -59.3 ± 1.6	-160.7 ± 5.1	-8.21 ± 0.07	58.1
T C A T C G A T G **C**	[d] -59.5 ± 1.2	-165.4 ± 2.3	-8.19 ± 0.28	50.7
T C G A T A T C G **C**	[c] -62.8 ± 1.2	-175.7 ± 3.9	-8.34 ± 0.12	50.7
C G C T A T A G C **T**	[d] -57.7 ± 2.4	-159.6 ± 5.7	-8.16 ± 0.59	51.0
C C G A T A T C G **T**	[c] -61.1 ± 1.8	-170.6 ± 5.5	-8.13 ± 0.13	50.0
T G C T A T A G C **C**	[d] -56.3 ± 1.3	-155.8 ± 3.1	-7.98 ± 0.34	50.2
T A G A G C T C T **C**	[c] -56.4 ± 2.0	-153.5 ± 6.3	-8.81 ± 0.12	55.2
C T C T C G A G A **T**	[d] -53.8 ± 1.8	-145.5 ± 4.3	-8.69 ± 0.52	55.4
C A G A G C T C T **T**	[c] -54.8 ± 2.1	-149.4 ± 6.4	-8.47 ± 0.17	53.7
T T C T C G A G A **C**	[d] -50.3 ± 1.2	-135.6 ± 2.8	-8.28 ± 0.36	53.9

Table S1: Continued.

	ΔH° (kcal / mol)	ΔS° (cal / mol K)	ΔG°_{37} (kcal / mol)	T_M [b] (°C)
G T G A G C T C A **A**	[c] -56.6 ± 2.8	-154.3 ± 8.4	-8.79 ± 0.17	55.1
A A C G C G A G T **G**	[d] -53.5 ± 1.6	-144.6 ± 3.6	-8.64 ± 0.44	55.2
A T G A G C T C A **G**	[c] -57.7 ± 1.9	-157.0 ± 5.6	-9.05 ± 0.15	56.3
G A C T C G A G T **A**	[d] -53.8 ± 1.6	-145.1 ± 3.6	-8.85 ± 0.45	56.5
G G T A G C T A C **A**	[c] -59.2 ± 4.2	-162.7 ± 12.9	-8.77 ± 0.16	54.1
A C A T C G A T G **G**	[d] -59.4 ± 1.7	-163.3 ± 4.1	-8.72 ± 0.44	53.8
A G T A G C T A C **G**	[c] -63.3 ± 3.4	-173.4 ± 10.2	-9.46 ± 0.20	56.8
G C A T C G A T G **A**	[d] -63.8 ± 1.8	-175.2 ± 4.5	-9.47 ± 0.45	56.6
G C G A T A T C G **A**	[c] -62.1 ± 1.3	-172.8 ± 3.8	-8.51 ± 0.10	51.8
A G C T A T A G C **G**	[d] -59.1 ± 1.1	-163.6 ± 2.6	-8.39 ± 0.27	51.9
A C G A T A T C G **G**	[c] -63.1 ± 2.6	-174.1 ± 7.9	-9.06 ± 0.15	54.6
G G C T A T A G C **A**	[d] -60.4 ± 1.6	-166.0 ± 3.7	-8.92 ± 0.40	54.6
G A G A G C T C T **A**	[c] -56.4 ± 1.9	-153.5 ± 5.6	-8.84 ± 0.15	55.4
A T C T C G A G A **G**	[d] -52.8 ± 1.3	-142.4 ± 3.1	-8.67 ± 0.38	55.6
A A G A G C T C T **G**	[c] -57.1 ± 2.6	-154.7 ± 7.6	-9.08 ± 0.20	56.7
G T C T C G A G A **A**	[d] -53.1 ± 2.2	-142.8 ± 6.7	-8.86 ± 0.10	56.8
T T G A G C T C A **G**	[c] -55.2 ± 1.5	-150.0 ± 4.5	-8.63 ± 0.07	54.5
G A C T C G A G T **T**	[d] -57.2 ± 0.8	-156.4 ± 1.8	-8.71 ± 0.21	54.3
G T G A G C T C A **T**	[c] -57.3 ± 0.7	-156.5 ± 2.3	-8.80 ± 0.05	54.9
T A C T C G A G T **G**	[d] -57.6 ± 1.6	-157.4 ± 3.7	-8.81 ± 0.42	54.8
T G T A G C T A C **G**	[c] -59.9 ± 1.8	-164.5 ± 5.5	-8.93 ± 0.08	54.8
G C A T C G A T G **T**	[d] -58.2 ± 2.0	-158.9 ± 4.8	-8.86 ± 0.53	55.0
G G T A G C T A C **T**	[c] -62.8 ± 3.6	-174.8 ± 11.0	-8.62 ± 0.17	52.3
T C A T C G A T G **G**	[d] -67.1 ± 3.8	-187.8 ± 9.4	-8.79 ± 0.85	52.1

Table S1: Continued.

	ΔH° (kcal / mol)	ΔS° (cal / mol K)	ΔG°_{37} (kcal / mol)	T_M [b] (°C)
T C G A T A T C G G	[c] -58.9 ± 2.8	-163.0 ± 8.9	-8.40 ± 0.06	52.0
G C C T A T A G C T	[d] -63.5 ± 0.6	-177.3 ± 1.6	-8.53 ± 0.15	51.6
G C G A T A T C G T	[c] -61.6 ± 3.7	-170.4 ± 11.5	-8.71 ± 0.14	53.1
T G C T A T A G C G	[d] -63.9 ± 0.9	-177.6 ± 2.3	-8.77 ± 0.22	52.8
T A G A G C T C T G	[c] -56.4 ± 1.5	-153.3 ± 4.5	-8.89 ± 0.06	55.7
G T C T C G A G A T	[d] -58.0 ± 0.6	-158.1 ± 1.9	-8.96 ± 0.03	55.6
G A G A G C T C T T	[c] -57.7 ± 1.6	-157.2 ± 4.7	-8.95 ± 0.10	55.7
T T C T C G A G A G	[d] -56.6 ± 1.9	-153.9 ± 4.6	-8.90 ± 0.52	55.7
Core sequences				
C G A T A T C G [e]	[c] -55.7 ± 3.9	-157.1 ± 12.1	-6.93 ± 0.12	44.1
G C T A T A G C	[d] -51.8 ± 0.6	-145.1 ± 1.4	-6.82 ± 0.15	44.0
G T A G C T A C [e]	[c] -55.1 ± 2.3	-155.0 ± 7.0	-7.04 ± 0.10	44.9
C A T C G A T G	[d] -51.4 ± 0.6	-143.3 ± 1.3	-6.95 ± 0.14	44.9
A G A G C T C T	[c] -49.5 ± 1.8	-134.5 ± 5.7	-7.76 ± 0.07	50.6
T C T C G A G A	[d] -50.1 ± 0.8	-136.7 ± 1.8	-7.76 ± 0.22	50.5
T G A G C T C A	[c] -50.7 ± 0.7	-138.4 ± 2.2	-7.73 ± 0.04	50.1
A C T C G A G T	[d] -50.3 ± 0.7	-137.3 ± 1.6	-7.72 ± 0.18	50.1

[a] The top strand of each duplex is represented in the 5' to 3' orientation and the bottom strand is shown in the 3' to 5' direction. Terminal mismatch nearest neighbors are represented in bold. Mismatches are underlined. [b] T_M calculated using 10^{-4} M total strand concentration.

[c] Thermodynamic parameters from averaging the fits of melting curves. Reported errors are standard deviations in the precision of the data. [d] Thermodynamic parameters from T_M^{-1} vs. $\ln(C_T)$ plots. Reported errors are standard deviations in the precision propagated from the slope and intercept of the $1/T_M$ vs. $\ln C_T$ plot. [e] Data from reference (19).

Table 1: Thermodynamic Parameters for Hairpin Oligomer Association and Oligomer Duplex Formation.

	ΔH° (kcal / mol)	ΔS° (cal / mol K)	ΔG°_{37} (kcal mol)	T_M (°C)
Systems with Elementary Interfaces[b]				
A A G C C T T G A - A C A A C G [c] (i) C G C G G A A C T / T G T T G C	-50.7 ± 4.1	-141.7 ± 11.3	-6.74 ± 0.27	49.3
A A G C C T T G T - T C A A C G [c] (ii) C G C G G A A C A / A G T T G C	-52.8 ± 4.2	-146.3 ± 11.7	-7.43 ± 0.30	53.1
G C A A C T - T G T T C C G A A [c] (iii) C G T T G A / A C A A G G C C C	-63.5 ± 5.1	-179.2 ± 14.3	-7.99 ± 0.32	53.2
A A G C C T T G A - T C A A C G [c] (iii) C G C G G A A C T / A G T T G C	-53.6 ± 4.3	-149.3 ± 11.9	-7.34 ± 0.29	52.3
A A G C C T T G T - A C A A C G [c] C G C G G A A C A / T G T T G C	-45.1 ± 3.6	-124.8 ± 10.0	-6.42 ± 0.26	48.4
A A G C C T T G C - A C A A C G [c] C G C G G A A C G / T G T T G C	-46.1 ± 3.7	-128.5 ± 10.3	-6.26 ± 0.25	46.9
A A G C C T T G T - G C A A C G [c] C G C G G A A C A / C G T T G C	-52.2 ± 4.2	-144.4 ± 11.5	-7.39 ± 0.30	53.1
A A G C C T T G G - T C A A C G [c] C G C G G A A C C / A G T T G C	-53.6 ± 4.3	-148.2 ± 11.9	-7.67 ± 0.31	54.4
A A G C C T T G A - C C A A C G [c] C G C G G A A C T / G G T T G C	-51.3 ± 4.1	-140.2 ± 11.2	-7.81 ± 0.31	56.2

Table 1: Continued.

	ΔH° (kcal / mol)	ΔS° (cal / mol K)	ΔG°_{37} (kcal mol)	T_M (°C)
5' A A G C C T T G C - T C A A C G 3' [c] 3' C G C G G A A C G / A G T T G C 5'	-46.1 ± 3.7	-126.3 ± 10.1	-6.90 ± 0.28	51.7
5' A A G C C T T G A - G C A A C G 3' [c] 3' C G C G G A A C T / C G T T G C 5'	-48.2 ± 3.9	-131.7 ± 10.5	-7.32 ± 0.29	53.9
5' G C A A C A - G G T T C C G A A 3' [c] 3' C G T T G T / C C A A G G C C 5'	-51.9 ± 4.2	-147.0 ± 11.8	-6.34 ± 0.25	46.3
5' A A G C C T T G G - A C A A C G 3' [c] 3' C G C G G A A C C / T G T T G C 5'	-50.4 ± 4.0	-139.5 ± 11.2	-7.10 ± 0.28	51.7
5' A A G C C T T G T - C C A A C G 3' [c] 3' C G C G G A A C A / G G T T G C 5'	-54.2 ± 4.3	-147.9 ± 11.8	-8.29 ± 0.33	58.3
5' A A G C C T T G C - G C A A C G 3' [c] 3' C G C G G A A C G / C G T T G C 5'	-47.6 ± 3.8	-130.9 ± 10.5	-6.96 ± 0.28	51.6
5' A A G C C T T G G - C C A A C G 3' [c] 3' C G C G G A A C C / G G T T G C 5'	-52.1 ± 4.2	-140.1 ± 11.2	-8.67 ± 0.35	61.9
5' A A G C C T T G C - C C A A C G 3' [c] 3' C G C G G A A C G / G G T T G C 5'	-53.3 ± 4.3	-146.5 ± 11.7	-7.92 ± 0.32	56.1
5' A A G C C T T G G - G C A A C G 3' [c] 3' C G C G G A A C C / C G T T G C 5'	-49.3 ± 3.9	-135.2 ± 10.8	-7.35 ± 0.29	53.8

Table1: Continued.

	ΔH° (kcal / mol)	ΔS° (cal / mol K)	ΔG°_{37} (kcal / mol)	T_M (°C)
Systems with Dangling Ends at the Interface[b]				
5' A A G C C T T G C - G C A A C G 3' [c] 3' C G C G G A A C G / C G T T G C 5' A	-44.4 ± 3.6	-121.4 ± 9.7	-6.76 ± 0.27	51.1
5' A A G C C T T G C - G C A A C G 3' [c] 3' C G C G G A A C G / C G T T G C 5' A	-48.0 ± 3.8	-131.8 ± 10.5	-7.16 ± 0.29	52.9
5' A A G C C T T G C - G C A A C G 3' [c] 3' C G C G G A A C G / C G T T G C 5' T	-46.3 ± 3.7	-127.4 ± 10.2	-6.83 ± 0.27	51.0
5' A A G C C T T G C - G C A A C G 3' [c] 3' C G C G G A A C G / C G T T G C 5' T	-49.0 ± 3.9	-136.6 ± 10.9	-6.59 ± 0.26	48.7
5' A A G C C T T G C - G C A A C G 3' [c] 3' C G C G G A A C G / C G T T G C 5' A A	-37.6 ± 3.0	-102.0 ± 8.2	-5.91 ± 0.24	46.4
5' A A G C C T T G C - G C A A C G 3' [c] 3' C G C G G A A C G / C G T T G C 5' T T	-36.2 ± 2.9	-97.3 ± 7.8	-6.03 ± 0.24	47.8
5' A A G C C T T G C - G C A A C G 3' [c] 3' C G C G G A A C G / C G T T G C 5' A T	-44.0 ± 3.5	-123.5 ± 9.9	-5.67 ± 0.23	43.0
5' A A G C C T T G C - G C A A C G 3' [c] 3' C G C G G A A C G / C G T T G C 5' T A	-43.6 ± 3.5	-119.5 ± 9.6	-6.53 ± 0.26	49.6
5' A A G C C T T G G - T C A A C G 3' [c] 3' C G C G G A A C C / A G T T G C 5' A	-47.2 ± 3.8	-129.8 ± 10.4	-6.90 ± 0.28	51.3

Table1: Continued.

	ΔH° (kcal / mol)	ΔS° (cal / mol K)	ΔG°_{37} (kcal / mol)	T_M (°C)
Systems with Extra Central Nucleotide at the Interface[b]				
A A G C C T T G C A G C A A C G [c] C G C G G A A C G / C G T T G C	-44.4 ± 3.6	-122.2 ± 9.8	-6.50 ± 0.26	49.2
A A G C C T T G T A C C A A C G [c] C G C G G A A C A / G G T T G C	-45.0 ± 3.6	-124.6 ± 10.0	-6.32 ± 0.25	47.7
Oligomers				
T C A A C G [c] A G T T G C	-38.5 ± 3.1	-108.3 ± 8.7	-4.94 ± 0.20	38.0
A C A A C G [c] T G T T G C	-36.1 ± 2.9	-99.3 ± 7.9	-5.32 ± 0.21	41.4
G C A A C G [c] C G T T G C	-42.5 ± 3.4	-117.1 ± 9.4	-6.21 ± 0.25	47.5
C C A A C G [c] G G T T G C	-38.6 ± 3.1	-106.2 ± 8.5	-5.69 ± 0.23	44.2
T G T T G C [c] A C A A C G	-37.1 ± 3.0	-101.1 ± 8.1	-5.79 ± 0.23	45.3
A G T T G C [c] T C A A C G	-37.0 ± 3.0	-101.0 ± 8.1	-5.70 ± 0.23	44.2

[a] T_M calculated using 10^{-4} total strand concentration.

[b] The top strand of each system is conventionally represented in the 5' to 3' orientation. Nucleotides involved in coaxial stacking interfaces are represented in bold.

[c] Parameters obtained by averaging the results of melt fit and TM^{-1} vs. $\ln(C_T/4)$ plot methods. Errors are estimated to be 8% for ΔH° and ΔS° and 4% for ΔG°_{37}.

(i), (ii), (iii) Melting curves for these systems are shown in Figure 2.

Table 2: Thermodynamic Parameters for Coaxial Stacking[a].

	ΔH°(coaxial stacking) (kcal / mol)	ΔS°(coaxial stacking) (cal / mol K)	ΔG°_{37}(coaxial stacking) (kcal / mol)
Elementary Interfaces[b]			
G A - A C C T / T G	-14.6 ± 5.0	-42.4 ± 13.8	-1.42 ± 0.34
G T - T C C A / A G	-14.3 ± 5.2	-38.0 ± 14.6	-2.49 ± 0.36
C T - T G G A / A C	-26.6 ± 5.9	-78.2 ± 16.5	-2.29 ± 0.39
G A - T C C T / A G	-15.1 ± 5.3	-41.0 ± 14.8	-2.40 ± 0.35
G T - A C C A / T G	-9.0 ± 4.6	-25.5 ± 12.8	-1.10 ± 0.33
G C - A C C G / T G	-10.0 ± 4.7	-29.2 ± 13.0	-0.94 ± 0.33
G T - G C C A / C G	-9.6 ± 5.4	-27.3 ± 14.9	-1.18 ± 0.39
G G - T C C C / A G	-15.1 ± 5.3	-39.9 ± 14.7	-2.73 ± 0.36
G A - C C C T / G G	-12.7 ± 5.1	-34.0 ± 14.1	-2.12 ± 0.39
G C - T C C G / A G	-7.6 ± 4.8	-18.0 ± 13.3	-1.97 ± 0.34
G A - G C C T / C G	-5.6 ± 5.1	-14.6 ± 14.1	-1.11 ± 0.38
C A - G G G T / C C	-14.8 ± 5.1	-45.9 ± 14.3	-0.56 ± 0.34
G G - A C C C / T G	-14.2 ± 5.0	-40.2 ± 13.7	-1.78 ± 0.35

Table 2: Continued.

	ΔH°(coaxial stacking) (kcal / mol)	ΔS°(coaxial stacking) (cal / mol K)	ΔG°_{37}(coaxial stacking) (kcal / mol)
G T - C C C A / G G	-15.6 ± 5.3	-41.8 ± 14.6	-2.61 ± 0.40
G C - G C C G / C G	-5.0 ± 5.1	-13.8 ± 14.1	-0.75 ± 0.37
G G - C C C C / G G	-13.5 ± 5.2	-33.9 ± 14.1	-2.98 ± 0.41
G C - C C C G / G G	-14.7 ± 5.3	-40.3 ± 14.5	-2.23 ± 0.39
G G - G C C C / C G	-6.8 ± 5.2	-18.1 ± 14.3	-1.14 ± 0.38
Interfaces with Dangling Ends[b]			
G C - G C C G / C G A	-1.9 ± 4.9	-4.3 ± 13.5	-0.55 ± 0.37
G C - G C C G / C G A	-5.5 ± 5.1	-14.7 ± 14.1	-0.95 ± 0.38
G C - G C C G / C G T	-3.8 ± 5.0	-10.3 ± 13.8	-0.62 ± 0.37
G C - G C C G / C G T	-6.4 ± 5.2	-19.5 ± 14.4	-0.38 ± 0.36
G C - G C C G / C G A A	5.0 ± 4.5	15.1 ± 12.4	0.30 ± 0.34

Table 2: Continued.

	ΔH°(coaxial stacking) (kcal / mol)	ΔS°(coaxial stacking) (cal / mol K)	ΔG°_{37}(coaxial stacking) (kcal / mol)
G C - G C **C G / C G** **T T**	6.3 ± 4.5	19.8 ± 12.2	0.18 ± 0.35
G C - G C **C G / C G** **A T**	-1.4 ± 4.9	-6.4 ± 13.6	0.55 ± 0.34
G C - G C **C G / C G** **T A** **A**	-1.1 ± 4.9	-2.4 ± 13.4	-0.32 ± 0.36
G G - T C **C C / A G** **A**	-8.6 ± 4.9	-21.5 ± 13.5	-1.96 ± 0.34
Interface with Extra Central Nucleotide[b]			
G C A G C **C G / C G**	-1.9 ± 4.9	-5.1 ± 13.5	-0.29 ± 0.36
G T A C C **C A / G G**	-6.4 ± 4.7	-18.5 ± 13.1	-0.64 ± 0.34

[a] These parameters and their corresponding errors are deduced from Table 1 as described in the text.
[b] The top strand of each duplex is conventionally represented in the 5' to 3' orientation. Nucleotides involved in coaxial stacking interfaces are represented in bold.

Table S1: Extinction coefficients of hairpins at 25 °C

	experimental[a] ($L\ mol^{-1}\ cm^{-1}$)	calculated[b] ($L\ mol^{-1}\ cm^{-1}$)
A A G C C T T G **G** T C A A C G C G C G G A A C **C**	188847	192310
A A G C C T T G **A** T C A A C G C G C G G A A C **T**	188718	195950
A A G C C T T G **C** T C A A C G C G C G G A A C **G**	186139	191610
A A G C C T T G **T** T C A A C G C G C G G A A C **A**	191071	195810
A A G C C T T G **A A** C A A C G C G C G G A A C **T**	194330	200950
A A G C C T T G **T A** C A A C G C G C G G A A C **A**	193953	200810
A A G C C T T G **G A** C A A C G C G C G G A A C **C**	188889	197310
A A G C C T T G **C A** C A A C G C G C G G A A C **G**	194623	196610
A A G C C T T G **A G** C A A C G C G C G G A A C **T**	192953	197350
A A G C C T T G **T G** C A A C G C G C G G A A C **A**	195968	197630
A A G C C T T G **G G** C A A C G C G C G G A A C **C**	195177	193710

Table S1: Continued

AAGCCTTGCGCAACG CGCGGAACG	190944	193010
AAGCCTTGACCAACG CGCGGAACT	192663	194350
AAGCCTTGTCCAACG CGCGGAACA	193532	194210
AAGCCTTGGCCAACG CGCGGAACC	195094	192390
AAGCCTTGCCCAACG CGCGGAACG	192864	190010
GCAACA-G TTCC AA CCAAGGCCC	200806	196010
GCAACT-T TTCC AA ACAAGGCCC	206650	192810
AAGCCTTGCAGCAACG CGCGGAACG	202688	206510
AAGCCTTGTACCAACG CGCGGAACA	204450	208010
AAGCCTTGCGCAACG CGCGGAACG A	191282	193010
AAGCCTTGCGCAACG CGCGGAACG T	198343	193010

[a] Calculated with Equation 3.
[b] Calculated with Equation 4.

Table S2. Thermodynamic Parameters for Hairpin Oligomer Association and Oligomer Duplex Formation.

		ΔH° (kcal / mol)	ΔS° (cal / mol K)	ΔG°_{37} (kcal / mol)	T_M[a] (°C)
Elementary interfaces[a]					
A A G C C T T G A - A C A A C G (i)	c	-48.1 ± 2.5	-133.2 ± 8.1	-6.8 ± 0.1	50.1
C G C G G A A C T / T G T T G C	d	-53.3 ± 2.7	-150.1 ± 8.7	-6.7 ± 0.0	48.4
A A G C C T T G T - T C A A C G (ii)	c	-55.7 ± 4.3	-155.8 ± 14.0	-7.4 ± 0.2	52.1
C G C G G A A C A / A G T T G C	d	-49.9 ± 1.7	-136.9 ± 5.6	-7.4 ± 0.0	54.2
A A G C C T T G A - T C A A C G (iii)	c	-51.9 ± 4.0	-143.6 ± 12.7	-7.3 ± 0.1	52.8
C G C G G A A C T / A G T T G C	d	-55.4 ± 2.1	-155.0 ± 4.9	-7.3 ± 0.5	51.7
G C A A C T - T G T T C C G A A	c	-61.6 ± 2.6	-173.1 ± 7.9	-8.0 ± 0.1	53.7
C G T T G A / A C A A G G C C C	d	-65.5 ± 3.1	-185.2 ± 9.8	-8.0 ± 0.0	52.8
A A G C C T T G T - A C A A C G	c	-48.0 ± 2.3	-134.4 ± 7.2	-6.4 ± 0.1	47.3
C G C G G A A C A / T G T T G C	d	-42.2 ± 1.9	-115.2 ± 6.3	-6.5 ± 0.1	49.5
A A G C C T T G C - A C A A C G	c	-45.9 ± 1.1	-127.9 ± 3.3	-6.3 ± 0.1	47.1
C G C G G A A C G / T G T T G C	d	-46.3 ± 4.0	-129.2 ± 13.1	-6.3 ± 0.1	46.8
A A G C C T T G T - G C A A C G	c	-54.7 ± 2.7	-152.4 ± 9.3	-7.4 ± 0.2	52.2
C G C G G A A C A / C G T T G C	d	-49.7 ± 3.4	-136.3 ± 11.0	-7.4 ± 0.1	53.9
A A G C C T T G G - T C A A C G	c	-53.5 ± 4.3	-147.7 ± 13.6	-7.7 ± 0.2	54.4
C G C G G A A C C / A G T T G C	d	-53.8 ± 3.4	-148.7 ± 11.1	-7.7 ± 0.1	54.3
A A G C C T T G A - C C A A C G	c	-52.9 ± 2.2	-145.4 ± 7.3	-7.8 ± 0.1	55.6
C G C G G A A C T / G G T T G C	d	-49.6 ± 2.4	-134.9 ± 5.5	-7.8 ± 0.7	56.8
A A G C C T T G C - T C A A C G	c	-48.6 ± 2.8	-134.6 ± 9.3	-6.9 ± 0.2	50.6
C G C G G A A C G / A G T T G C	d	-43.6 ± 2.2	-118.1 ± 7.3	-6.9 ± 0.1	52.8
A A G C C T T G A - G C A A C G	c	-49.7 ± 3.5	-136.6 ± 11.4	-7.3 ± 0.1	53.4
C G C G G A A C T / C G T T G C	d	-46.6 ± 1.8	-128.8 ± 5.7	-7.3 ± 0.0	49.9

Table S2: Continued.

		ΔH° (kcal / mol)	ΔS° (cal / mol K)	ΔG°_{3}- (kcal mol)	T (°
G C A A C A - G G T T C C G A A C G T T G T / C C A A G G C C C	[c]	-53.1 ± 3.8	-150.8 ± 13.0	-6.3 ± 0.2	4
	[d]	-50.8 ± 3.3	-143.3 ± 10.9	-6.4 ± 0.1	4
A A G C C T T G G - A C A A C G C G C G G A A C C / T G T T G C	[c]	-48.4 ± 3.1	-133.1 ± 10.4	-7.1 ± 0.1	5
	[d]	-52.3 ± 2.0	-145.9 ± 6.5	-7.1 ± 0.0	5
A A G C C T T G T - C C A A C G C G C G G A A C A / G G T T G C	[c]	-58.0 ± 2.1	-160.2 ± 6.4	-8.4 ± 0.2	5
	[d]	-50.3 ± 2.1	-135.7 ± 4.7	-8.2 ± 0.6	5
A A G C C T T G C - G C A A C G C G C G G A A C G / C G T T G C	[c]	-47.0 ± 3.9	-129.0 ± 12.9	-7.0 ± 0.1	5
	[d]	-48.2 ± 0.9	-132.8 ± 2.9	-7.0 ± 0.0	5
A A G C C T T G G - C C A A C G C G C G G A A C C / G G T T G C	[c]	-57.3 ± 5.4	-156.4 ± 17.0	-8.8 ± 0.3	6
	[d]	-46.9 ± 2.7	-123.9 ± 5.8	-8.5 ± 0.9	6
A A G C C T T G C - C C A A C G C G C G G A A C G / G G T T G C	[c]	-55.3 ± 2.6	-152.7 ± 8.1	-7.9 ± 0.1	5
	[d]	-51.4 ± 1.5	-140.2 ± 3.4	-7.9 ± 0.5	5
A A G C C T T G G - G C A A C G C G C G G A A C C / C G T T G C	[c]	-45.9 ± 1.0	-124.4 ± 3.5	-7.3 ± 0.1	5
	[d]	-52.6 ± 3.5	-146.0 ± 11.3	-7.4 ± 0.1	5
Interfaces with dangling ends[a]					
A A G C C T T G C - G C A A C G C G C G G A A C G / C G T T G C A	[c]	-45.2 ± 2.9	-123.8 ± 9.4	-6.8 ± 0.2	5
	[d]	-43.7 ± 3.4	-119.1 ± 11.1	-6.8 ± 0.1	51.5
A A G C C T T G C - G C A A C G C G C G G A A C G / C G T T G C T	[c]	-46.7 ± 3.7	-128.7 ± 11.8	-6.8 ± 0.1	50.9
	[d]	-46.0 ± 3.3	-126.2 ± 7.5	-6.8 ± 1.0	5
A A G C C T T G C - G C A A C G C G C G G A A C G / C G T T G C A	[c]	-48.6 ± 3.0	-133.7 ± 9.4	-7.2 ± 0.1	5
	[d]	-47.4 ± 3.2	-129.8 ± 10.5	-7.2 ± 0.1	5

Table S2: Continued.

		ΔH° (kcal / mol)	ΔS° (cal / mol K)	ΔG°_{37} (kcal / mol)	T_M[a] (°C)
5'-A A G C C T T G C - G C A A C G-3'	[c]	-46.3 ± 2.6	-127.9 ± 8.7	-6.6 ± 0.2	43.6
3'-C G C G G A A C G / C G T T G C-5'	[d]	-51.6 ± 3.7	-145.4 ± 12.1	-6.5 ± 0.1	42.3
T					
5'-A A G C C T T G G - T C A A C G-3'	[c]	-49.2 ± 3.5	-136.6 ± 11.0	-6.9 ± 0.2	44.8
3'-C G C G G A A C C / A G T T G C-5'	[d]	-45.1 ± 2.9	-123.1 ± 9.4	-6.9 ± 0.1	45.7
A					
5'-A A G C C T T G C - G C A A C G-3'	[c]	-39.7 ± 4.6	-109.1 ± 15.2	-5.8 ± 0.2	38.2
3'-C G C G G A A C G / C G T T G C-5'	[d]	-35.5 ± 2.9	-95.0 ± 5.9	-6.0 ± 1.1	39.8
A A					
5'-A A G C C T T G C - G C A A C G-3'	[c]	-44.7 ± 3.3	-126.0 ± 11.0	-5.6 ± 0.2	36.6
3'-C G C G G A A C G / C G T T G C-5'	[d]	-43.2 ± 3.7	-121.0 ± 12.2	-5.7 ± 0.2	37.2
A T					
5'-A A G C C T T G C - G C A A C G-3'	[c]	-38.5 ± 3.0	-105.1 ± 10.0	-6.0 ± 0.2	39.3
3'-C G C G G A A C G / C G T T G C-5'	[d]	-33.9 ± 1.3	-89.6 ± 4.5	-6.1 ± 0.1	41.0
T T					
5'-A A G C C T T G C - G C A A C G-3'	[c]	-43.3 ± 9.0	-118.5 ± 30.1	-6.5 ± 0.3	43.1
3'-C G C G G A A C G / C G T T G C-5'	[d]	-43.9 ± 1.0	-120.5 ± 3.4	-6.5 ± 0.0	43.2
T A					
Interfaces with extra central nucleotide					
5'-A A G C C T T G C A G C A A C G-3'	[c]	-43.8 ± 9.4	-120.2 ± 31.0	-6.5 ± 0.2	43.0
3'-C G C G G A A C G / C G T T G C-5'	[d]	-45.0 ± 3.5	-124.1 ± 11.5	-6.5 ± 0.1	42.7
5'-A A G C C T T G T A C C A A C G-3'	[c]	-45.0 ± 2.8	-124.5 ± 9.1	-6.3 ± 0.2	41.6
3'-C G C G G A A C A / G G T T G C-5'	[d]	-45.0 ± 4.0	-124.8 ± 13.4	-6.3 ± 0.2	41.4

Table S2: Continued.

		ΔH° (kcal / mol)	ΔS° (cal / mol K)	ΔG°_{37} (kcal / mol)	T_M[a] (°C)
Oligomers					
	T C A A C G	[c] -40.6 ± 2.5	-115.4 ± 9.1	-4.8 ± 0.3	30.5
	A G T T G C	[d] -36.5 ± 1.9	-101.2 ± 6.3	-5.1 ± 0.1	31.9
	A C A A C G	[c] -38.3 ± 2.0	-106.8 ± 6.5	-5.2 ± 0.2	33.4
	T G T T G C	[d] -33.9 ± 1.6	-91.8 ± 5.3	-5.4 ± 0.1	34.6
	G C A A C G	[c] -43.9 ± 2.4	-121.6 ± 7.4	-6.2 ± 0.1	40.7
	C G T T G C	[d] -41.1 ± 2.1	-112.6 ± 6.9	-6.2 ± 0.1	41.2
	C C A A C G	[c] -41.0 ± 2.6	-114.1 ± 8.9	-5.6 ± 0.2	36.4
	G G T T G C	[d] -36.2 ± 0.9	-98.2 ± 1.8	-5.8 ± 0.3	37.8
	T G T T G C	[c] -38.8 ± 3.8	-106.6 ± 12.1	-5.7 ± 0.1	37.5
	A C A A C G	[d] -35.5 ± 2.8	-95.7 ± 9.2	-5.8 ± 0.1	38.4
	A G T T G C	[c] -37.1 ± 2.9	-101.2 ± 9.4	-5.7 ± 0.1	36.9
	T C A A C G	[d] -36.9 ± 2.4	-100.8 ± 8.0	-5.7 ± 0.1	36.9

[a] T_M calculated for 4×10^{-4} total strand concentration.

[b] The top strand of each system is conventionally represented in the 5' to 3' orientation. Nucleotides involved in coaxial stacking interfaces are represented in bold.

[c] Parameters obtained from averaging fits of melting curves. Reported errors are standard deviations in the precision of the fitted data.

[d] Parameters obtained from T_M^{-1} vs. $\ln(C_T/4)$ plots. Reported errors are standard deviations in the precision propagated from the slope and intercept of the $1/T_M$ vs. $\ln (C_T/4)$ plot.

[(i), (ii), (iii)] $1/T_M$ vs. $\ln(C_T/4)$ plots for these systems are shown in Figure S1.